NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

**TRUST IN AUTOMATED SYSTEMS:
THE EFFECT OF AUTOMATION LEVEL ON
TRUST CALIBRATION**

by

James C. Walliser

June 2011

Thesis Advisor: Lawrence G. Shattuck
Second Reader Robert L. Shearer

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE		Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 2011	Master's Thesis

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
Trust in Automated Systems: The Effect of Automation Level on Trust Calibration	
6. AUTHOR(S) Capt. James C. Walliser	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number NPS.2011.0061.EP7-A.

12a. DISTRIBUTION / AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
Approved for public release; distribution is unlimited	A

13. ABSTRACT (maximum 200 words)

Automated systems perform functions that were previously executed by a human. When using automation, the role of the human changes from operator to supervisor. For effective operation, the human must appropriately calibrate trust in the automated system. Improper trust leads to misuse and disuse of the system. The responsibilities of an automated system can be described by its level of automation. This study examined the effect of varying levels of automation and accuracy on trust calibration.

Thirty participants were divided into three groups based on the system's level of automation and provided with an automated identification system. Within the Virtual Battlespace 2 environment, participants controlled the video feed of an unmanned aircraft while they identified friendly and enemy personnel on the ground. Results indicate a significant difference in the ability to correctly identify targets between levels of automation and accuracy. Participants exhibited better calibration at the management by consent level of automation and at the lower accuracy level. These findings demonstrate the necessity of continued research in the field of automation trust.

14. SUBJECT TERMS trust, calibration, automation, levels of automation	15. NUMBER OF PAGES 85
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500 Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

TRUST IN AUTOMATED SYSTEMS: THE EFFECT OF AUTOMATION LEVEL ON TRUST CALIBRATION

James C. Walliser
Captain, United States Air Force
B.S., United States Air Force Academy, 2005

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN HUMAN SYSTEMS INTEGRATION

from the

NAVAL POSTGRADUATE SCHOOL
June 2011

Author: James C. Walliser

Approved by: Lawrence G. Shattuck
Thesis Advisor

Robert L. Shearer
Second Reader

Robert F. Dell
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Automated systems perform functions that were previously executed by a human. When using automation, the role of the human changes from operator to supervisor. For effective operation, the human must appropriately calibrate trust in the automated system. Improper trust leads to misuse and disuse of the system. The responsibilities of an automated system can be described by its level of automation. This study examined the effect of varying levels of automation and accuracy on trust calibration.

Thirty participants were divided into three groups based on the system's level of automation and provided with an automated identification system. Within the Virtual Battlespace 2 environment, participants controlled the video feed of an unmanned aircraft while they identified friendly and enemy personnel on the ground. Results indicate a significant difference in the ability to correctly identify targets between levels of automation and accuracy. Participants exhibited better calibration at the management by consent level of automation and at the lower accuracy level. These findings demonstrate the necessity of continued research in the field of automation trust.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

The employment of automated systems is expanding across the modern battlefield. The growth of automation has not eliminated the human from the system; it has transformed the human's role. With the trend toward increasing automation, human roles have changed from operators to supervisors. This change does not necessarily mean that human workload has been reduced. Instead, cognitive resources are applied to different tasks, such as anticipating the automation and understanding the actions of the automation. Nonetheless, highly automated systems will be critical for the supervision of multiple unmanned systems across the battlefield. The changing but continuous role that humans maintain with automated systems requires an understanding of the human-automation relationship. One aspect of this relationship that had yet to be explored was the process by which humans calibrate trust in automated systems.

This study examined the calibration of trust at three levels of automation and two levels of accuracy. The levels of automation were decision support, management by consent, and management by exception. Accuracy of the automation was set at 75% and 90%. The experiment was a mixed design in which level of automation was a between subjects factor while accuracy was a within subjects factor. The experiment was conducted in the Human Systems Integration Laboratory at Naval Postgraduate School using Virtual Battlespace 2 software. Thirty participants were tasked to identify enemy and friendly targets as the operator of a video feed from an unmanned aircraft. In support of the task, participants were given the assistance of an automated identification system. Participants were divided into three groups and provided with information about the responsibilities of the automation. The descriptions corresponded to one of three levels of automation listed above.

The results of this study suggest that a system's level of automation may influence an operator's ability to calibrate trust. There was a statistically significant difference in the correct identification percentage between levels of

automation. When informed that a system was automated at the management by consent level, participants outperformed groups at the decision support and management by exception levels. Better performance may indicate better trust calibration, but not in the direction hypothesized. The accuracy of the automated system also influenced the correct identification percentage. Performance was better at the 90% accuracy level but a greater percentage of automation errors were identified at the 75% accuracy level. The difference was statistically significant. We hypothesized that trust calibration would decrease as accuracy decreased, but trust calibration appeared to increase as accuracy decreased.

This study explored the process of trust calibration in automated systems. New automated systems are fielded regularly and the level of automation should be carefully considered early in system development. An understanding of the cognitive processes in play on a human-automation team is vital to the future integration of highly automated systems onto the battlefield. We must examine the manner in which humans build trust in automated systems and how trust relates to effective operation. The goal of future research should not be to divide the tasks between humans and machines; the efforts need to focus on how humans and machines work together.

ACKNOWLEDGMENTS

I thank my thesis adviser, Dr. Lawrence Shattuck, and second reader, Lieutenant Colonel Robert Shearer, for their guidance and vision throughout the thesis process. Specifically, I thank Dr. Shattuck for providing me direction while I was struggling to find a topic. My thesis would not have been successful without his insightful questions and reasoned analysis. I thank Lieutenant Colonel Shearer for his constant support through the ups and downs of the data analysis and his tireless effort to provide the answers we needed.

I also thank Dr. Nita Shattuck for her advice while we worked through the experimental design. Thanks to Dr. Kip Smith for the exceptional counsel at every stage in the process.

Most importantly, I thank my wife, Nora. Her love and encouragement have kept me going throughout my time at Naval Postgraduate School. I truly could not have accomplished so much without her by my side.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. PROBLEM STATEMENT

Automated systems are widely used in civilian and military applications. Examples can be as simple as the turn-by-turn directions available in a car's global positioning system or as complex as the flight controls of unmanned aircraft systems (UAS). According to Lee and See (2004), automation is technology that actively selects data, transforms information, makes decisions or controls processes. The United States military has recognized the value of automation on the battlefield across the spectrum of military operations. By October 2006, coalition UASs had logged nearly 400,000 flight hours and unmanned ground vehicles had responded to more than 11,000 improvised explosive device situations (Department of Defense [DoD], 2007). An unmanned vehicle is a powered vehicle that does not carry a human operator, can be operated autonomously or remotely, can be expendable or recoverable and can carry a lethal or nonlethal payload (DoD, 2007). The need for unmanned systems continues to grow. Each year, Combatant Commanders submit an integrated priorities list of their theater's capability gaps. In 2008, 17 of the top 99 prioritized gaps could have been addressed by unmanned systems, including 2 of the top 10 (DoD, 2007).

According to the United States Air Force (USAF) *UAS Flight Plan 2009–2047* (Department of the Air Force [DAF], 2009a), "Unmanned aircraft systems are one of the most "in demand" capabilities the USAF provides to the joint task force." Part of the USAF vision in this document was to harness automated systems to maximize Joint Force combat capabilities. One of ten key assumptions in the *UAS Flight Plan* was that automation is vital to increasing effects and cutting costs. As technologies advance, automated systems will compress the time to observe, orient, decide, and act. This sequence of activities is commonly referred to as the OODA Loop. A UAS will interpret the situation and act with little or no human interaction. As automation becomes more

sophisticated the need for an operator will be reduced. However, automation does not replace humans in the system; it changes the nature of the task. Rather than operating a UAS, the human will supervise its actions.

In a supervisory control environment, automation can be designed with varying levels of autonomy. In a highly automated system the computer makes all decisions and acts on its own. A minimally automated system might simply present the supervisor with options and defer to the supervisor to make the decision. Several classification systems to describe levels of automation have been proposed. Table 1 is the earliest description of levels of automation, developed by Sheridan and Verplank (1978).

Table 1. Levels of Automation (From Sheridan & Verplank, 1978)

Automation Level	Automation Description
1	The computer offers no assistance: human must make all decisions & actions
2	The computer offers a complete set of decision/action alternatives, or
3	narrows the selection down to a few, or
4	suggests one alternative, and
5	executes that suggestion if the human approves, or
6	allows the human a restricted time to veto before automatic execution, or
7	executes automatically, then necessarily informs humans, and
8	informs the human only if asked, or
9	informs the human only if it, the computer, decides to.
10	The computer decides everything and acts autonomously, ignoring the human

The Combatant Commanders and Military Departments identified precision target location and designation as the number two capability need to be filled by UASs (DoD, 2007). Target identification and designation capability can

be supported with varying levels of automation. A system with Level 3 automation might select the most likely targets for the supervisor to designate, while Level 6 automation would designate a target on its own, but give the supervisor an opportunity to veto the decision.

Employing an automated target designator has some inherent risk; targets must be identified with a high degree of certainty. Unfortunately, a highly automated system is not equivalent to a highly accurate system. There are numerous examples of the failure to correctly use automation in the military. The *USS Vincennes* disaster in 1978, and the destruction of a British Tornado and American F/A-18 in 2004 with a Patriot missile system, are two examples (Fisher & Kingma, 2001; 32nd Army Air and Missile Defense Command, 2003).

Highly automated decision aids are useful for rigid tasks, but they are not suited for decisions in dynamic environments. In dynamic situations, the automation may not be programmed to adapt, leading to a catastrophic failure for which the human supervisor is not prepared. For example, the *DoD UAS Roadmap* (2007) identifies the challenge of developing automation that considers rules of engagement. Not every situation is black and white; the supervisor must be capable of stepping in when the automation fails to interpret the gray areas. In automated systems, inability to adapt to novel situations is known as the "brittleness problem" (Guerlain & Bullemer, 1996; Guerlain, 1995).

Automation is often sold as a solution to reduce operator workload and enhance situational awareness. The USAF views any implementation of automation in the near future as a tool to decrease workload (*DAF*, 2009a). The assumption is that human and automation teams perform better than a human alone. Some research has shown that automation can improve performance in rigid situations requiring little flexibility in decision making (Endsley & Kaber, 1999). However, other research has shown that human operators often make errors through misuse or disuse of automated aids (Parasuraman & Riley, 1997). Misuse occurs when an operator over relies on an automated aid. The authors cite the crash of Eastern Flight 401 into the Florida Everglades as an example of

misuse. The crew did not realize the autopilot had disengaged and failed to monitor altitude, allowing the airliner to crash. Disuse is the under reliance on automation. Whenever an individual ignores an alarm, the system is being disused.

Simply improving the reliability of an automated aid will not lead to appropriate automation reliance or increase the overall performance of the human-automation team. Sorkin and Woods (1985) demonstrated that a more reliable automated aid did not lead to the best overall performance. The supervisor and automation work as a team and the operator must determine the appropriate circumstances to rely on automation. One of the factors known to influence a supervisor's decision to rely on automation is trust. Trust is the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability (Lee & See, 2004). However, trust alone is not enough to ensure appropriate reliance. An optimally automated system with high detection rates and low false alarm rates may seem more trustworthy, but the supervisor must know when to trust the system and when not to trust the system. Inappropriate levels of trust result in disuse or misuse of an automated aid. Overreliance or misuse occurs when the supervisor trusts an automated aid that is less reliable than manual operation. Distrusting an automated aid that is more reliable than manual operation leads to under-reliance or disuse.

Since errors in the level of trust in self—and trust in automation—lead to misuse and disuse, it is important that supervisors appropriately calibrate their trust. Calibration is the correspondence between a person's perception of the reliability of an agent and the true reliability of that agent (van Dongen & van Maanen, 2006). In order to achieve the best performance, supervisors of unmanned systems must be capable of calibrating trust in the system. Appropriate calibration results in decreased misuse and disuse of automated systems.

Accurate calibration is crucial in a military environment where operators must make quick decisions while relying on the guidance of automated aids. We must understand the process by which operators calibrate their trust in automation. Though the concept of trust in automation has been thoroughly researched, there are some topics remaining to be addressed. For one, researchers have yet to examine how trust is calibrated with specific levels of automation. The current study investigated how levels of automation impact an operator's ability to calibrate trust in the system.

B. OBJECTIVES

This research explored how operator performance was affected by a system's level of automation and accuracy. Specifically, this study:

- Assessed the ability of the human operator to calibrate trust at varying levels of automation
- Assessed the ability of the human operator to calibrate trust at varying automation accuracy levels

C. RESEARCH QUESTIONS

- How do we measure trust calibration?
- Is the ability to accurately calibrate trust associated with level of automation?
- Is the ability to accurately calibrate trust associated with automation accuracy?

D. HUMAN SYSTEMS INTEGRATION (HSI)

The human plays a central role in every weapons system. Manned or unmanned, there will always be some type of interaction between the system and the human. The Naval Postgraduate School (2010) describes HSI as follows:

> Human Systems Integration (HSI) acknowledges that the human is a critical component in any complex system. It is an interdisciplinary

> approach that makes explicit the underlying tradeoffs across the HSI domains, facilitating optimization of total system performance.

By following the principles of HSI, practitioners can optimize total system performance across the system's lifecycle. Simply recognizing the role of the human in the system is just the first step toward performance optimization. The integration of the human into the system must be approached from multiple domains. The HSI domains are:

- Human Factors Engineering
- Human Survivability
- Health Hazards
- System Safety
- Habitability
- Manpower
- Personnel
- Training

The true HSI process occurs when the HSI practitioner defines the human requirements in each of these domains and considers the tradeoffs that must occur. Tradeoffs among the HSI domains create a balance among cost, schedule and technical performance parameters. This is how total system performance is optimized.

Three of the HSI domains are particularly relevant to the present study; human factors engineering, training, and personnel. Human factors engineering "involves the understanding and comprehensive integration of human capabilities into system design" (DAF, 2009b). Human factors engineers integrate cognitive, physical, sensory, social capabilities to create human-systems interfaces in support of operation, maintenance, support and sustainment. According to Wickens, Lee, Liu, and Becker (2004), the goal of human factors is to design

systems that enhance performance, increase safety, and increase user satisfaction. In the context of the present study, an understanding of the human's interaction with automation can influence the design of interfaces that support optimal performance. The ability to accurately calibrate trust could decrease the disuse and misuse of automated systems. Proper use of the system is directly related to performance, safety, and user satisfaction.

Training, "encompasses the instruction and resources required to provide personnel with the requisite knowledge, skills, and abilities to properly operate, maintain, and support systems" (DAF, 2009b). According to the Defense Acquisition University, training program design uses "analyses, methods, and tools to ensure systems training requirements are fully addressed and documented by systems designers and developers to achieve a level of individual and team proficiency that is required to successfully accomplish tasks and missions" (Defense Acquisition University, 2011). The HSI practitioner must determine who needs instruction, what to teach them, and how to provide the instruction. The results of the present study can influence training program design for unmanned vehicle operators. Understanding the complexities of trust calibration can change the who, what, and how of training program design.

HSI practitioners working in the personnel domain consider the "total human characteristics and skill requirements for a system to support full operational capabilities necessary to operate, maintain, and support a system" (Defense Acquisition University, 2011). The knowledge, skills, abilities and aptitudes translate directly to personnel requirements and methods to recruit, test and select personnel. The personnel selected for a system influence training, manpower, and design requirements. The results of the present study will support HSI practitioners attempting to define knowledge, skill, ability, and aptitude requirements for unmanned vehicle operators.

E. THESIS ORGANIZATION

This thesis is divided into five chapters. Chapter II provides a review of the relevant literature regarding levels of automation, trust, and calibration. Chapter III describes the method used to conduct the experiment. The discussion includes details regarding participants, materials, variables, and procedures. Chapter IV is a report of the results of the experiment. The thesis ends with a discussion of conclusions that can be drawn from the results, as well as recommendations for follow-on research.

II. LITERATURE REVIEW

A. OVERVIEW

This literature review is divided into three sections. It begins with a discussion of levels of automation. The second section describes the concept of trust in automation. The third section concludes the literature review with an examination of the process of trust calibration.

B. LEVELS OF AUTOMATION

Unmanned systems have the capacity to execute actions that previously required a human to complete. Functions now allocated to unmanned systems include tasks that humans do not wish to perform or cannot perform as accurately or reliably (Parasuraman, Sheridan, & Wickens, 2000). Researchers have devoted a great deal of effort to determine just what tasks machines perform more accurately and reliably. The challenge of allocating functions to humans and machines has led researchers to ask, “What can humans do better than machines?” The earliest attempt to divide responsibility between humans and machines resulted in Fitts’ List (Fitts, 1951). The list, also known as MABA-MABA, describes tasks that men are better at performing and tasks that machines are better at performing. Fitts’ List is shown in Table 2.

Table 2. Fitts' List (From Fitts, 1951)

Men Are Better At
Detection of small amounts of sensory information.
Perception of patterns.
Improvisation and flexibility of procedures.
Exercising judgment.
Recall of relevant information at the appropriate moment.
Performing inductive reasoning.
Machines Are Better At
Rapid response to data.
Application of great force smoothly and precisely.
Executing repetitive, routine functions.
Performing deductive reasoning and computations.
Brief storage of information for immediate use.
Execution of multiple tasks simultaneously.

Fitts' List served as a starting point to describe the division of labor between humans and machines. However, the list does not describe the division of responsibility when humans and machines work together. Humans rely on automated systems to perform tasks that machines are better at performing. A device that accomplishes a function that was previously carried out by a human operator is an automated system (Parasuraman et al., 2000). While automation is required to replace tasks previously performed by humans, the capabilities and responsibilities of automation are not equal. Automation is not an all or none quality; tasks may be partially or completely controlled by the automated system. Highly automated systems are granted greater control over tasks and require less human interaction.

Simply describing a system as highly automated does not convey sufficient information about the capabilities of the automation. The construct of levels of automation was developed to serve as a more precise description of an automated system's capabilities and the requirements of the human operator. The earliest description of the levels of automation appeared in a technical report to the Office of Naval Research written by Sheridan and Verplank (1978). The authors developed a model to describe the division of responsibility between humans and automation. High level of automation corresponded with greater automation responsibility. Ten levels of automation were defined. Level 1, a completely manual task, had the least automation and Level 10, a completely computer controlled task, had the most automation.

Sheridan and Verplank's model for human automation interaction was followed by several similar models. Rouse and Rouse (1983) proposed three levels of automation to describe the human-automation relationship. Manual control, Sheridan and Verplank's first level, was described as dormant automation. The system remains inactive unless initiated by the operator. Management-by-consent was equivalent to Level 5 in Sheridan and Verplank's model. At this level, automation proposes action but cannot act without approval by the operator. The third level, management-by-exception, was parallel to Sheridan and Verplank's Level 6 in which automation will act unless explicitly directed not to by the operator. While the levels described by Rouse and Rouse could be mapped to similar descriptions in Sheridan and Verplank's model there were several gaps. Notably, Rouse and Rouse did not include Levels 2, 3, and 4 in which the automation provides suggestions to the operator. In addition, they excluded Levels 7–10 in which the automation acts without input from the operator.

Endsley (1987) developed a level of automation hierarchy to specifically describe expert decision aid systems. Many aspects of Sheridan and Verplank's model were incorporated into the new hierarchy. Notably, Levels 2, 3, and 4 were combined into a single level called "Decision Support." This removed the

distinction between multiple alternatives, few alternatives and a single decision option. In addition, the author removed levels seven, eight, and nine from the earlier model, eliminating the description of how the system provides feedback to the operator. The changes made by Endsley resulted in a five level hierarchy:

1. Manual Control—No assistance from the system
2. Decision Support—Operator receives recommendations from the system
3. Consensual Artificial Intelligence—System performs task if operator consents
4. Monitored Artificial Intelligence—System performs task unless operator vetoes
5. Full Automation—No operator interaction

A more recent scale of human automation interaction was presented by Sheridan (2002, p 53) as a simplification of the original model. The model still retains the original format, describing the role of the computer and the human at each level. However, two levels have been removed from the earlier model. Levels 2 and 3 were combined to do away with the distinction between a system that presents a complete list of alternatives and a system that provides a narrow selection. The second change was the removal of Level 9, the level at which the system informs the operator only if the automation deems it necessary. The remainder of the scale of degrees of automation was unchanged. The revised scale of degrees of automation is depicted in Table 3.

Table 3. Revised Scale of Degrees of Automation (From Sheridan, 2002)

	A Scale of Degrees of Automation
1.	The computer offers no assistance: human must do it all.
2.	The computer suggests alternative ways to do the task
3.	The computer suggests one way to do the task AND
4.	…executes that suggestion if the human approves, OR
5.	…allows the human a restricted time to veto before automatic execution, OR
6.	…executes automatically, then necessarily informs the human, OR
7.	…executes automatically, and then informs the human only if asked.
8.	The computer selects the method, executes the task, and ignores the human.

Careful examination of the various scales of automation reveals similarities. For one, each scale includes a management by consent and management by exception level. Management by consent is the level of automation at which the system suggests a solution and will act on that solution with the consent of the human (Rouse & Rouse, 1983). On Sheridan's (2002) scale, it falls under Level 4 and is also referred to as consensual artificial intelligence by Endsley (1987). Management by exception is the level of automation at which the system selects a solution and allows the human time to veto the action before it is automatically executed (Rouse & Rouse, 1983). Management by exception corresponds with Level 5 on Sheridan's (2002) scale; Endsley (1987) calls this monitored artificial intelligence. There is also some agreement about the level that Endsley calls decision support. Both Sheridan's (2002) scale of degrees of automation and Sheridan and Verplank's (1978) levels of automation include a level at which the automation simply provides recommendations for the operator. However, Rouse and Rouse (1983) did not include this level in their description.

Level of automation describes the division of responsibility between a human and an automated system. However, the level of automation does not describe how operators use and perceive the system. Regardless of the level of automation, operators must develop trust in the system. Mistrusted automated systems will used inappropriately and ineffectively.

C. TRUST IN AUTOMATION

Trust is a critical component in the interaction between humans and machines. Interest in trust began as an examination of human-human trust. In this context, trust has been defined as "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another" (Rousseau, Sitkin, Burt & Camerer, 1998). While this definition is accepted, trust has been difficult to define with a single statement. Barber (1983) described trust as a combination of three expectations: 1) natural and moral laws will persist; 2) those we interact with are technically competent; and 3) those we interact with will carry out their fiduciary responsibility. Persistence, competency, and fiduciary responsibility together are the building blocks of trust. We must also consider the dynamics of trust, or how our perceptions of persistence, competency, and fiduciary responsibility are developed. Rempel et al. (1985) proposed that trust is a dynamic expectation and follows a developmental sequence based on predictability, dependability and faith. Early in a relationship trust is based on predictability, which is a result of consistent behavior. As time goes on dependability will become the dominant factor in trust. Specific behaviors become less important than a stable disposition in times of risk or vulnerability. The final stage of trust development is faith. Individuals look at past predictability and dependability then form expectations of behavior for future situations.

Human-automation researchers saw trust as a significant factor influencing overall performance. When attempting to define trust in automation, researchers drew parallels from sociological studies of human-human

interactions. Just like interpersonal relationships, humans and machines form a dyad in which trust is a significant factor of performance. Muir (1987) was among the first to consider trust in the man-machine domain. She developed a two-dimensional framework for studying trust that was a cross of the components of trust and the dynamic formation of trust (Table 4). Muir viewed persistence, competency, and responsibility as the most complete characterization of the components of trust, while trust was formed by predictability, dependability and faith. Muir (1994) proposed a model for trust in human-automation relationships in which humans compare their perceptions of persistence, competence and responsibility with their expectations. The product of the comparison between perceived performance and expected performance is trust.

Table 4. Basis of expectation (From Muir, 1987)

	Basis of Expectation At Different Levels of Experience		
Expectation	Predictability (of acts)	Dependability (of dispositions)	Faith (in motives)
Persistence			
Natural Physical	Events conform to natural laws	Nature is lawful	Natural laws are constant
Natural Biological	Human life has survived	Human survival is lawful	Human life will survive
Moral Social	Humans and computers act 'decent'	Humans and computers are 'good' and 'decent' by nature	Humans and computers will continue to be 'good' and 'decent' in the future
Technical Competence	j's behavior is predictable	j has a dependable nature	j will continue to be dependable in the future
Fiduciary Responsibility	j's behavior is consistently responsible	j has a responsible nature	j will continue to be responsible in the future

One of the earliest studies of trust in a human-machine system was performed by Lee and Moray (1992), as an extension of a study performed by Muir (1989). The purpose of the experiment was to develop a better understanding of human machine-trust. In this experiment, participants were asked to balance safety and performance while in control of a simulated pasteurization plant. The operators could vary control of the system from manual control, automatic control or mixed control throughout the experiment. A total of 60 trials were performed by each participant and they completed a trust questionnaire after each trial. The questions established the operator's subjective feelings about predictability, dependability, and faith in the automation. The results of this experiment suggested that reliance on an automated aid is not simply dependent on the perceived trustworthiness of the aid, but also the operator's self confidence.

Research has demonstrated that human-automation teams may exhibit less than optimal performance (Parasuraman & Riley, 1997). Human-only teams also demonstrate suboptimal performance. In the sociology community this poor performance is referred to as process loss. Poor performance has been attributed to cognitive, motivational, and social factors (Mullen, Johnson, & Salas, 1991). Dzindolet, Pierce, Beck, and Dawe (1999) applied the aspects of process loss to human-machine teams to develop a broad model of automation use. While Muir (1994) considered only cognitive factors, Dzindolet et al. (1999) incorporated cognitive, motivational, and social factors. The manner in which humans process information from an automated aid is the cognitive process. Mosier and Skitka (1996) have defined reliance on an automated aid in a heuristic manner as automation bias. The human is subject to motivational processes as part of a human-machine team. In a sociological context, responsibility for the end product is diffused amongst team members. This may lead to the human on a human-automation team feeling reduced responsibility for the task and inappropriate reliance on the automated aid. The concept of social process built on the work of Lee and Moray (1992, 1994) when they found the

decision to rely on automation is dependent on trust in the aid and operator self confidence. The comparison between perceived reliability of the aid and perceived reliability of manual control is known as perceived utility. The perceived utility of the automated aid along with automation bias directly contribute to the operator's relative trust. The complete model for automation use is depicted in Figure 1.

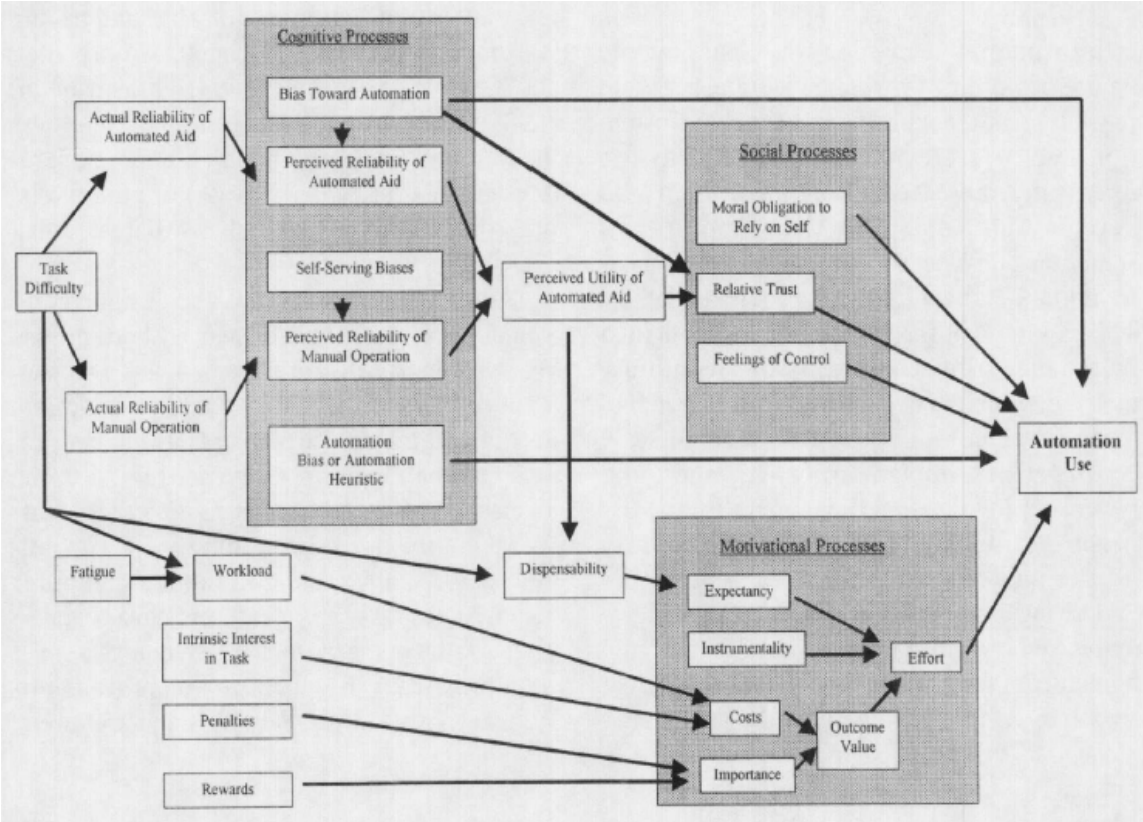


Figure 1. Model for Automation Use (From Dzindolet et al., 1999)

Dzindolet et al. (1999) performed numerous experiments to test the propositions of their new model. Over the course of several studies, participants were presented with photographs of Fort Sill, Oklahoma and asked to perform a visual detection task. Half of the images contained soldiers in camouflage and participants identified the soldier in images while using an "automated" contrast detection aid with varying levels of accuracy. Cognitive processes were controlled by presenting the automation's decision after the operator's decision

had been made. Motivational processes were controlled by setting the level of effort required to use the automation equal to the effort without automation. The authors measured participants' decision to rely on automation. Over the course of four experiments, the authors supported the prediction that automation use is determined by perceived utility of the aid, which is a result of a comparison between operator and aid performance.

Reliance on a system is not simply dependent on trust. Operators decide when to rely on an automated system through a comparison of their own perceived reliability and the perceived reliability of the automated aid. This comparison leads to the development of relative trust, which feeds into automation use. The process of developing relative trust in a system is referred to as trust calibration.

D. TRUST CALIBRATION

The comparison between trust in self and trust in automation is closely related to calibration of trust. Muir (1987) describes calibration as the user setting his or her trust level in correspondence with the machine's trustworthiness and using the machine accordingly. The properly calibrated operator knows when to rely on the automated system (e.g., appropriate trust) and when to rely on manual control (e.g., appropriate distrust) (Muir, 1994). The improperly calibrated operator exhibits false trust when relying on poor automation and false distrust when discounting good automation. In the context of Dzindolet, Pierce, Beck and Dawe's (2001) framework to predict automation use, calibration falls between the cognitive and social processes; it is the perceived utility of the automated aid.

Lee and Moray (1992) encountered evidence of trust calibration in the course of their simulation of a pasteurization factory. Participants in the experiment were tasked as operators of the factory to optimize system output by controlling several processes. Operators were allowed to adopt manual control, automated control, or mixed control strategies. As the experiment progressed, the simulation was programmed to produce errors resulting in reduced system

performance. For the majority of the experiment, operators demonstrated a tendency to rely more heavily on manual control. Unexpectedly, chronic faults led to increased reliance on the automated aid even though trust in the automated aid decreased. Operator's confidence in their own abilities decreased as well. The authors concluded trust was a factor in automation reliance, but self confidence in manual control abilities also contributed.

Calibration is critical to the appropriate reliance on automated systems. Inappropriate reliance on automated aids is synonymous with misuse or disuse of that aid (Parasuraman & Riley, 1997). Misuse of an automated aid occurs when the operator incorrectly relies on automated control over manual control. In this situation, perceived utility of the automated aid is too high. High perceived utility stems from inflated perception of automation reliability and/or deflated perception of operator reliability. Disuse occurs when perception of automation reliability is deflated and/or perception of operator reliability is inflated. As a result, the operator may incorrectly rely on manual control over automated control.

Another study examined how operators estimate their own reliability and the reliability of decision aids (Dongen & Maanen, 2006). The authors hypothesized that underestimation of decision aid reliability is more prevalent than underestimation of self reliability. In the course of the experiment participants were asked to estimate reliability of a decision aid and self reliability on a prediction task involving a sequence of numbers. The results of the experiment support the hypothesis that underestimation of the decision aid was more prevalent than self underestimation. In addition, they found that under-trust in own performance decreased over time, while under-trust of the decision aid persisted.

The perceived reliability of an automated aid is should be dependent on the aid's actual reliability. One framework for automation utilization assumes that operators initially expect automated aids to perform at near perfect rates, referred to as automation bias. As a result, errors by automation are particularly salient to

operators; this leads them to underestimate system reliability (Dzindolet et al., 2001). Wiegmann, Rich, & Zhang (2001) performed an experiment to examine the relationship between actual and perceived reliability. The participants were presented with automated diagnostic aids of varying reliability. The three conditions were 60% reliable increasing to 80% reliable, constant 80% reliability, and 100% reliable decreasing to 80% reliable. Participants were asked to estimate the reliability of the aid at the completion of the trials. Results suggested that operators of automated decision aids are sensitive to changing levels of aid reliability. In addition, this study supported the framework of cognitive and social processes affecting automation use developed by Dzindolet et al. (2001). Perceived utility of the automation was lower than automation reliance, suggesting the interference of social processes.

One approach to measuring trust calibration is grounded in the principles of signal detection theory (Tanner & Swets, 1954). Generally, signal detection theory is a comparison between the true state and the perceived state. The operator's response results in a hit, miss, false alarm, or correct rejection. When applying signal detection theory to trust calibration there are three components to consider: the true state of the environment, the recommendation of the automation, and the response of the decision maker. In this scenario, the decision maker is unable to fully perceive the true state of the environment, or there is uncertainty in his perceptions. The automated system aids the operator by providing an interpretation of the state of the environment. Figure 2 depicts the relationship between ground truth, automation, and the decision maker. The human must make a choice by performing a comparison between the information provided by the automation and his own perceptions.

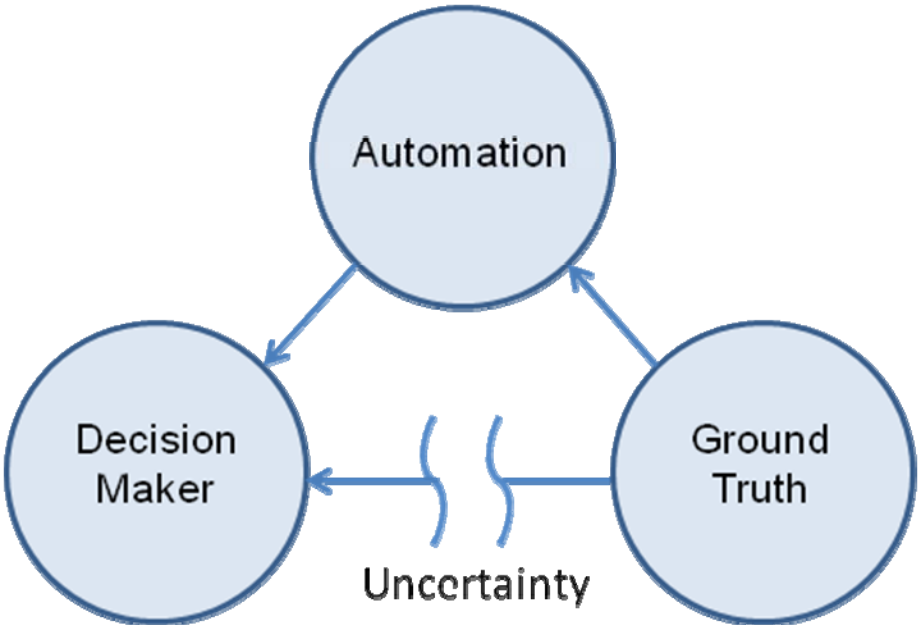


Figure 2. Relationship between ground truth, automation, and decision maker

Hits and correct rejections can occur when the operator appropriately agrees with or rejects the guidance of the automated aid. False alarms and misses are the result of misuse or disuse of the automated aid. An operator that is perfectly calibrated will know when to agree with the automated aid and when to disagree. Proper calibration is indicated by a high degree of hits and correct rejections. On the other hand, a high degree of misses and false alarms indicates poor calibration. The properly calibrated operator of an automated system should use the system as a tool to perceive the true state of the world, knowing when to accept or reject the system's indications and achieve a high rate of hits and correct rejections.

E. PRESENT STUDY

Designers of unmanned systems grapple with questions about level of automation and accuracy today. From an HSI standpoint, knowledge about the human-automation relationship can influence human factors, training, and personnel. The present study examines the effects of changes to the level of automation and accuracy of automated systems. It is hypothesized that

increasing level of automation will decrease the operator's ability to calibrate trust. In addition, decreasing accuracy of the automation will decrease the operator's ability to calibrate trust. This section describes the manner in which these hypotheses were derived.

A few researchers have explored the relationship between trust and levels of automation. Ruff, Narayanan, and Draper (2002) measured operator trust and correct detection of decision aid failure when controlling unmanned aircraft at varying levels of automation. The authors selected the Rouse and Rouse (1983) levels of automation: manual control, management by consent, and management by exception. In this experiment, operators were tasked to control one to four unmanned aircraft in a virtual environment as they searched for and engaged four ground targets. The automation provided decision aiding to the operator for changes in system state. At the completion of the experiment, participants were asked to rate their trust in the automated system, using subjective ratings based on the work of Masalonis and Parasuraman (1999). Results indicated that even a 5% error rate led to a significant drop in trust at higher automation levels. In addition, correct rejections were significantly lower at the management-by-exception level and this level consistently received the lowest trust ratings. The authors recommend that high levels of automation do not necessarily result in better performance. In some situations, optimal performance may be achieved with lower levels of automation.

Levels of automation are an effective way to describe the division of responsibilities in a human-automation team. Researchers have proposed differing automation level classification systems. Ruff et al. (2002) selected Rouse and Rouse's (1983) description of the levels of automation for their study. Similarities and differences among the classification systems were discussed earlier. The present study considers the similarities among classification systems and uses three levels of automation for examination: decision support, management by consent, and management by exception.

The study by Ruff et al. (2002) explored trust in levels of automation and the occurrence of correct detections of automation aid failure, but an area that has not been fully examined is the process of calibrating trust in automation. In the study by Ruff et al. (2002), correct detection of automation failures was recorded as a total for the entire trial. However, the process of calibration occurs over time, the operator must be provided feedback at regular intervals in order to refine his/her perception of the automation's reliability. In the present study, calibration is measured using the framework of signal detection theory. If calibration improves over time, it will be indicated by an increased degree of hits and correct rejections.

The accuracy of the automation impacts the operator's perception of reliability. Ruff et al. (2002) used two levels of accuracy in their study, 100% and 95%. They found that just that small change led to decreased trust in the automation. One thing they could not measure was how changing accuracy effected the operator's ability to detect errors and calibrate trust. The present study varied automation accuracy to assess the impact on trust calibration.

THIS PAGE INTENTIONALLY LEFT BLANK

III. METHOD

A. METHOD OVERVIEW

The experiment consisted of a series of target detection tasks at varying levels of automation and accuracy. In the scenario, participants acted as the observer of a video feed from an unmanned aircraft. Their task was to identify enemy and friendly personnel as the unmanned aircraft flew along a scripted flight path. Each participant completed three trial runs containing 50 enemy targets and 50 friendly targets.

The study incorporated a mixed design. Participants were randomly assigned to one of three experimental groups (between subjects; decision support, management by consent, or management by exception). Each group experienced two levels of automation accuracy (within subjects; 75% and 90%).

After completion of a manual control trial with no automated guidance, the participants were informed that they would be aided in subsequent trials by an automated identification system. The description of the automated system was consistent with a specific level of automation. In reality, the automated identification system did not exist. Instead, targets were identified by the experimenters as part of the scripted scenario. The same targets were identified as enemy and friendly at each automation level. Accuracy of the automation was either 75% or 90%, meaning 25% or 10% of the indications were misses and false alarms.

B. PARTICIPANTS

1. Selection

The Naval Postgraduate School Institutional Review Board reviewed and approved the design of this study, in accordance with Department of the Navy and American Psychological Association standards. All participants provided informed consent by signing a form that notified them of their rights as

participants in the experiment. Participants were solicited through e-mail communication and personal contact. The study used a convenience sample taken from the Naval Postgraduate School population.

2. Demographic Make-up

Thirty participants (average age = 30.83, SD = 4.25 years) completed this study, 21 were male and nine were female. Participants were drawn from every United States military branch. In addition, several foreign military officers and civilians participated in the study. See Figure 3 to view the distribution of participants.

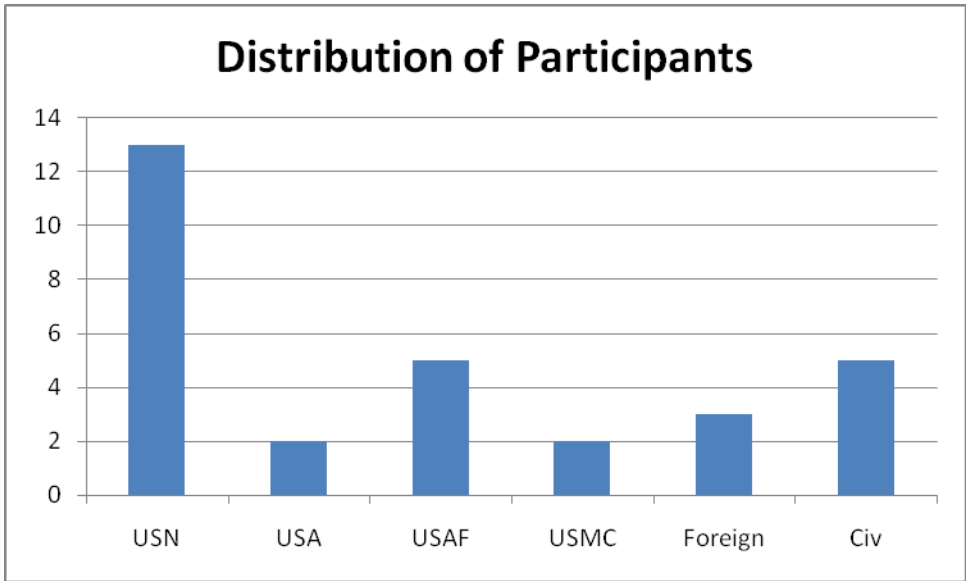


Figure 3. Distribution of Participants

Participants who were members of the military provided their time in service. Figure 4 depicts total years of military service, including enlisted time.

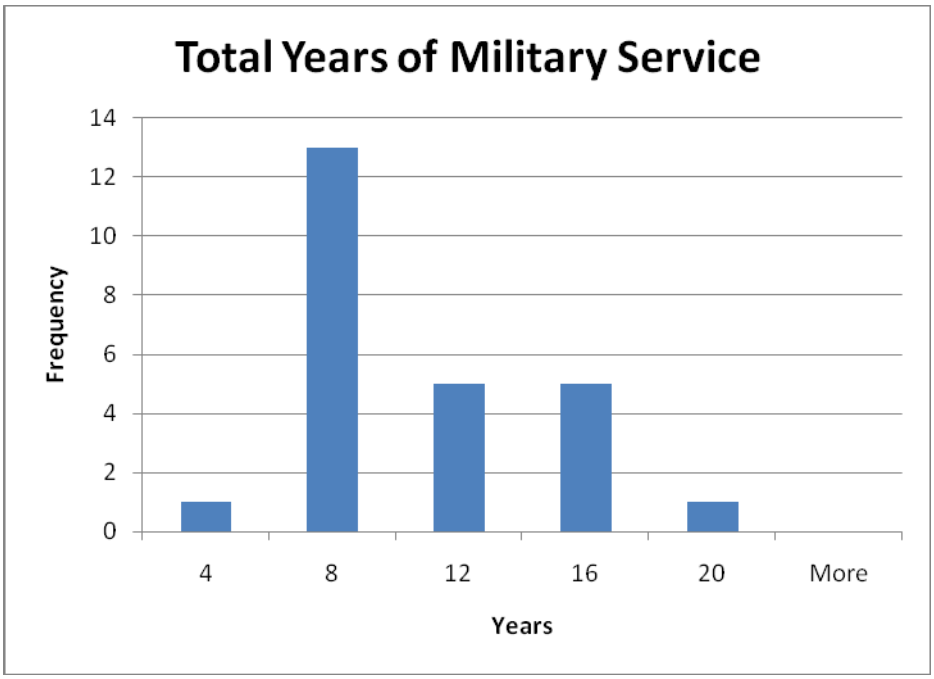


Figure 4. Total Years of Military Service

C. MATERIALS

1. Virtual Battle Space 2

Virtual Battlespace 2 is software that provides battlefield simulations. It was specifically designed by Bohemia Interactive for federal, state, and local government agencies. Primary uses of the software include training of doctrine, tactics, techniques, and procedures for squad and platoon operations.

2. Equipment

The experiment was run on a Dell Precision M6300 laptop computer with the following specifications:

- Attached Monitor: 24 inch Dell Flat Panel LCD Display
- Operating System: Microsoft Windows XP
- Processor: Intel Core 2 Duo T9500 @ 2.60 GHz
- 777 MHz, 3.5 GB of RAM

D. VARIABLES

1. Independent Variables

a. Level of Automation

Group A was informed that the automated identification system functioned at the level of automation corresponding to decision support. The automation identified potential friendly and enemy targets, and the operator used the information to make a decision.

Group B was informed that the automated identification system functioned at the level of automation corresponding to management by consent. The automation identified friendly and enemy targets and was prepared to act with consent of the operator.

Group C was informed that the automated identification system functioned at the level of automation corresponding to management by exception. The automation identified friendly and enemy targets and would act unless the operator vetoed.

b. Accuracy

The automated identification system operated at two levels of accuracy, Level 1 was set at 75% and Level 2 was set at 90%. Accuracy describes the number of friendly and enemy targets correctly identified. This was a within subjects variable, each participant completed a trial at 75% and 90% accuracy. Accuracy was counterbalanced so that half of the participants in each group experienced 75% accuracy first and half of the participants experienced 90% accuracy first.

2. Dependent Variables

There were three dependent variables: calibration, perceived reliability, and perceived utility. The number of friendly and enemy targets correctly identified by the participant measured calibration. Perceived reliability was

measured by responses to post-trial questionnaire Item 1. Perceived utility was measure by responses to post-trial questionnaire Item 2. See Appendix A for a copy of the post trial questionnaire.

E. PROCEDURE

Participants signed up for an hour-long experimental session as their schedules allowed. A single participant was tested during each session. The researcher distributed participants among the groups (A1, A2, B1, B2, C1, & C2) in the order they volunteered. The first participant was placed in Group A1, the second in A2, the third in B1, and so on. Participants met the researcher in the Human Systems Integration Laboratory. Upon completion of the Informed Consent documentation participants answered a demographic questionnaire (Appendix B).

Next, participants were provided with the initial scenario description. The description contained instructions for the participant, images of the targets, and the evaluation method. Participants were asked to summarize the directions to ensure accurate understanding of the task. The instructions read:

> There has been a recent increase in terrorist activities along a road of strategic importance. Little is known about the position of enemy and friendly personnel along the roadway. In order to collect intelligence on the disposition of forces an unmanned aircraft has been directed to scout the roadway. You are the operator of the unmanned aircraft video feed. The aircraft will fly a programmed route over the roadway while you manipulate the camera. When you encounter an enemy along the route press the key labeled E to indicate an enemy, the target will be destroyed when the aircraft has passed. When you encounter friendlies along the route, press the key labeled “F” to keep them safe from engagement.

The targets were placed along the flight path of the unmanned aircraft by the researchers. Three unique flight paths were programmed in Virtual Battlespace 2. An experimental trial was concluded when the aircraft flew a complete flight path. Participants completed three experimental trials without repeating flight paths. The Virtual Battlespace 2 mission editor provided a list of

targets from which the enemy and friendly personnel were chosen. The targets were selected to make accurate identification a challenge for the participants. Images of the targets are shown in Figure 5.

Figure 5. Enemy (left) and Friendly (right) targets.

The scenario description also included an explanation of the evaluation method. Performance was assessed by the number of hits, misses, false alarms, and correct rejections. Participants were provided with the following definitions:

- Hit—Identify an ENEMY target as an ENEMY
- Miss—Identify an ENEMY target as a FRIENDLY
- False Alarm—Identify a FRIENDLY target as an ENEMY
- Correct Rejection—Identify a FRIENDLY target as a FRIENLDY

Participants completed a practice trial for task familiarization. The practice trial was divided into five blocks, each containing 10 targets to be identified. During the practice trials, participants were informed that the first two targets of each block were always friendly and the second two were always enemy. When a participant used the keyboard to identify a target, an identification marker was recorded on a two-dimensional video map. The target markers and target location were combined to determine hits, misses, false alarms and correct rejections. At the end of each block, the researcher assessed performance with the participant by reviewing the map.

All participants began the experiment with the manual control trial. Participants were directed to identify targets, but they were not given the aid of an automated identification system. The trial contained 100 targets divided into five blocks of 20 targets. Each block contained an equal number of friendly and enemy targets. At the completion of a block the simulation was paused and the participant received feedback about the number of hits, correct rejections, false alarms, and misses.

Upon completion of the manual control scenario participants were given a new set of instructions. The instructions informed the participants that the unmanned aircraft had been upgraded with an automated identification system. The description of the automated identification system was dependent on the participant's experimental group. See Appendix C for the complete instructions given to the experimental groups. During the second and third trials, targets were identified by red arrows to indicate enemies and blue arrows to indicate friendlies. The participants completed one trial at 75% automation accuracy and a second trial at 90% automation accuracy. Each trial contained 100 targets, divided into five blocks of 20 targets. Each block contained an equal number of friendly and enemy targets. At the completion of a block the simulation was paused and the participant received feedback about the number of hits, correct rejections, false alarms, and misses.

At the completion of the second trial, participants were asked two questions. Responses to these questions were used to determine perceived reliability and perceived utility of the automated system. The questions were repeated at the end of the third trial as well. Perceived reliability was assessed by the following question:

> In the previous scenario, you were presented with 100 targets. Please estimate the percentage of times the automation was correct in its identification of individuals?

Perceived utility was assessed by the second question, which was presented as follows:

> If you were asked to scout an additional roadway would you prefer to identify and report targeting information without the use of the automated system or would you prefer to allow the automated system to identify and report targeting information without human supervision?

IV. RESULTS

A. CALIBRATION

The present study collected target identification data from participants that performed an identification task. The number of hits, misses, false alarms, and correct rejections were determined by comparison between the true target type and identified target type. Analysis was performed to determine the effect of automation accuracy, type of target, and level of automation on ability to correctly identify targets. An alpha level of 0.05 was used for all statistical tests.

1. Level of Automation

The present study examined correct identification percentage at three levels of automation. Participants were placed into one of three groups and experienced only a single level of automation. The number of correctly identified targets divided by the total number of targets was referred to as correct identification percentage (CIP). The mean CIP at LOA 1, decision support, was 88.2% (*SD*=18.5). Participants in LOA 2, management by consent, achieved a mean CIP of 96.0% (*SD*=5.8). The mean CIP at LOA 3, management by exception, was 89.9% (*SD*=9.8). Figure 6 presents a chart of mean correct identification percentage by sector and level of automation.

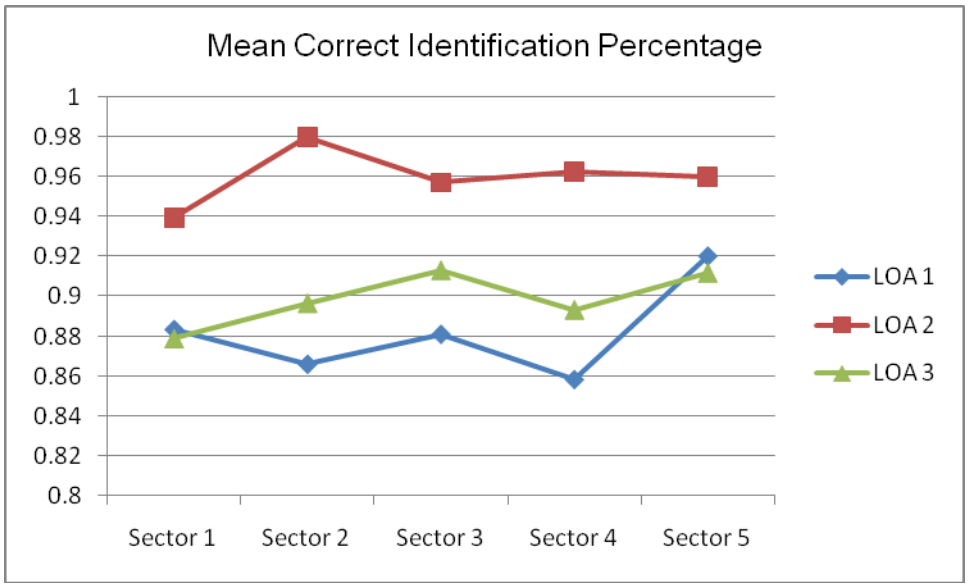


Figure 6. Mean Correct Identification Percentage

A one-way ANOVA was performed between CIP and LOA. A significant difference was detected between groups, $F(2, 297)=10.84$, $p<.0001$. Analysis of residuals indicated subject number 2 may have been an outlier. A Kruskal Wallis test confirmed the results of the one-way ANOVA, $x^2(2)=27.48$, $p<.0001$. The p-value of the ANOVA and Kruskal Wallis were the same, so the one-way ANOVA results were retained. Post hoc analysis of the results indicated that CIP at LOA 2 was greater than LOA 1 and LOA 3.

The receiver operator characteristics (ROC) of participants was also analyzed. Participants' hits, misses, false alarms, and correct rejections were grouped by LOA. Sensitivity (d') was calculated within each group and the results were placed on a scatter plot indicating the ratio of false alarms to hits. High sensitivity values indicate a high number of hits in relation to false alarms. Figure 7 depicts the ROC for each level of automation. Within a level of automation, sensitivity was calculated at each identification block. Figure 8 depicts the sensitivity across identification blocks at each LOA.

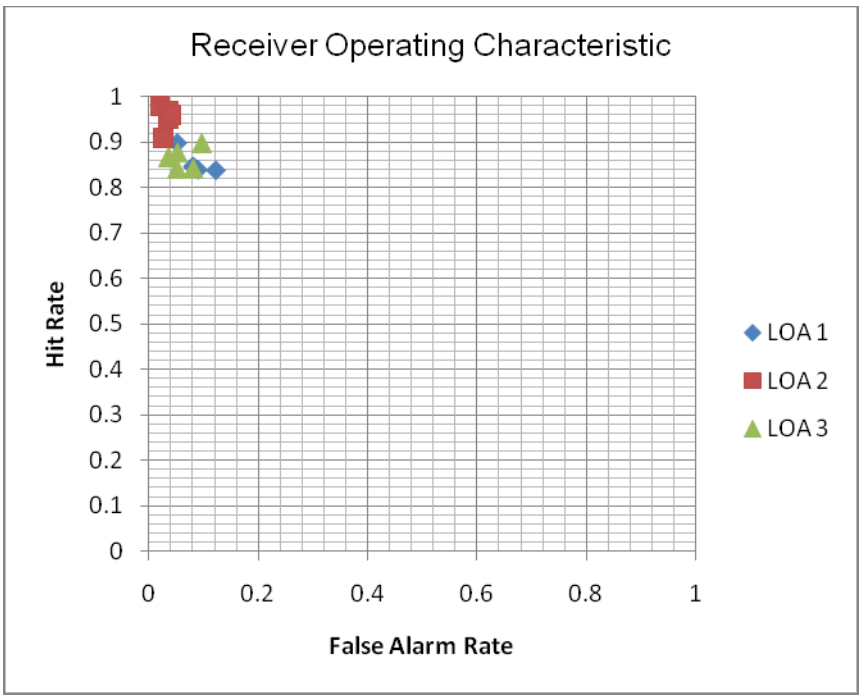


Figure 7. Receiver Operating Characteristic

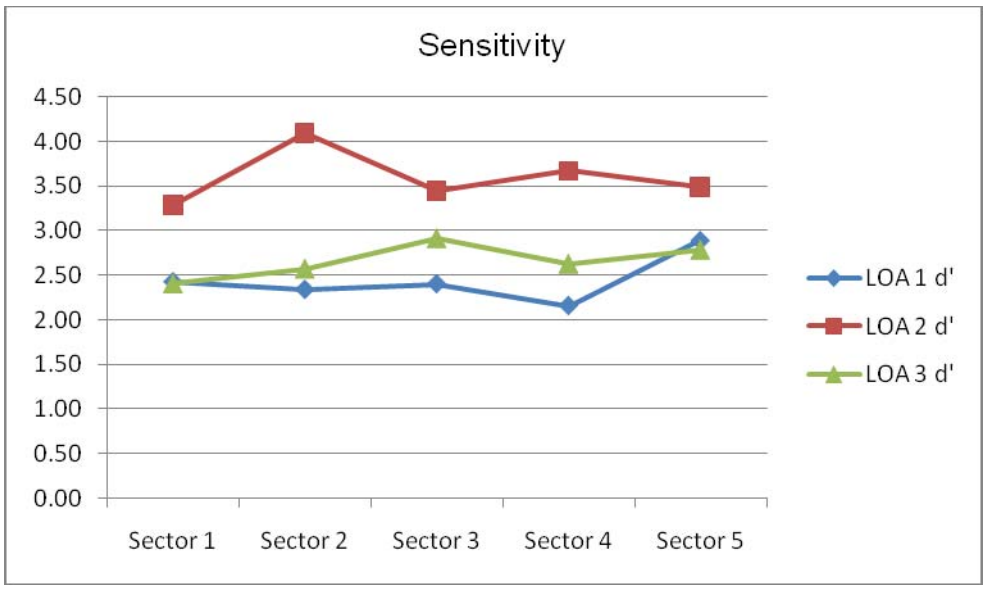


Figure 8. Sensitivity

A final chart was created by plotting CIP variance across zones at each level of automation. Variance at LOA 1, LOA 2, and LOA 3 (*M*=.041, .006, .016) did not appear to decrease across identification blocks. The plot of variance by

level of automation is depicted in Figure 9. Initially, the variance plots at each LOA did not overlap. However, two participants in LOA 1 performed far below the average. When their data were removed, the LOA 1 plot was very similar to the LOA 3 plot. The mean variance at LOA 1 dropped to 0.013. The adjusted plot of variance by level of automation is depicted in Figure 10.

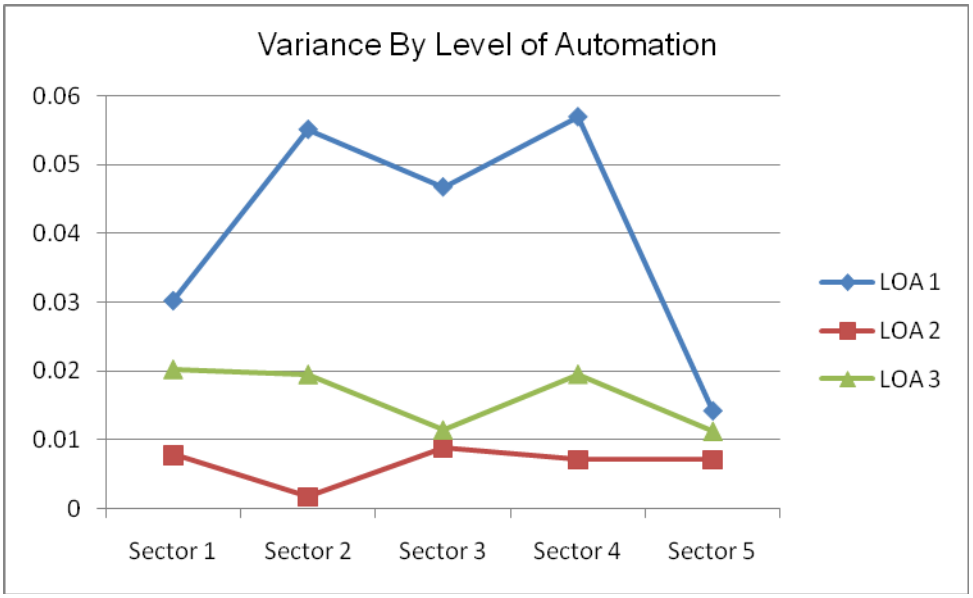


Figure 9. Variance By Level of Automation

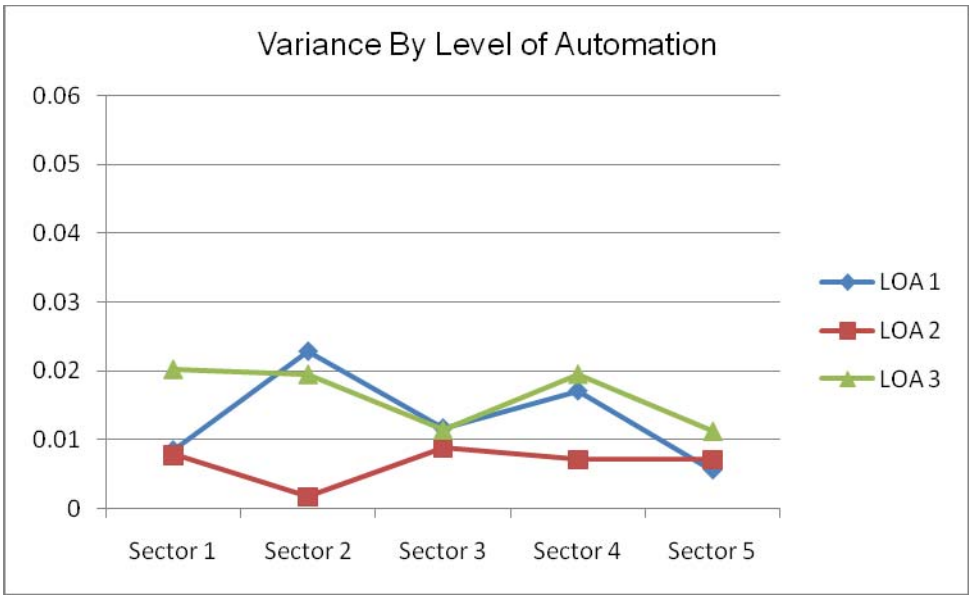


Figure 10. Adjusted Variance By Level of Automation

2. Automation Accuracy

Participants were presented with two levels of automation accuracy. One trial was completed at 75% accuracy and a second trial was completed at 90% accuracy. In the 75% accuracy condition, the mean CIP was 90.5% (*SD*=14.8). In the 90% accuracy condition, the mean CIP was 92.1% (*SD*=14.9). Over the course of one trial, participants identified targets in five sectors completed sequentially. Mean CIP was calculated for each sector and the values were plotted on a line chart. A complete experimental trial was represented by five connected data points, one for each sector in the trial. Figure 11 presents a chart of mean CIP by sector and accuracy level.

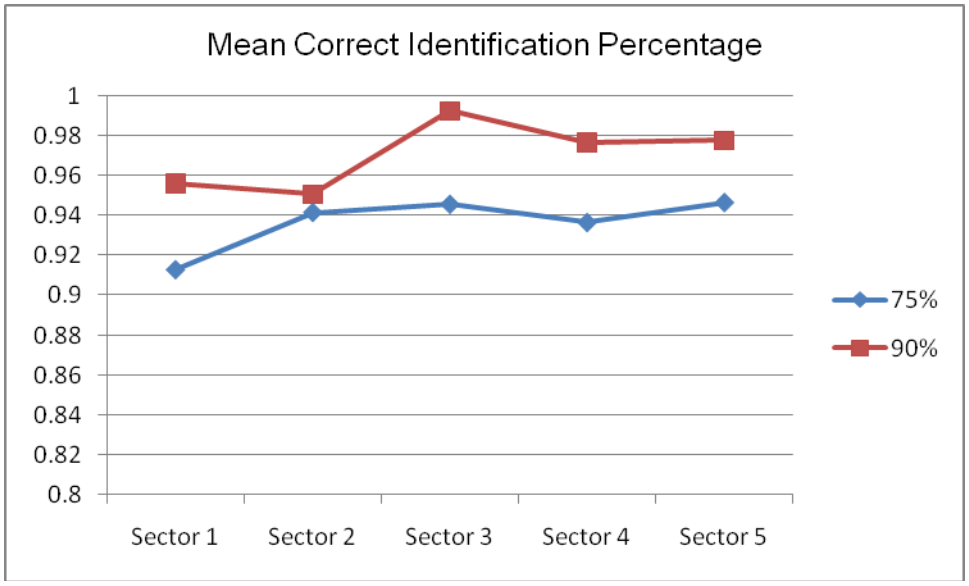


Figure 11. Mean Correct Identification Percentage

A repeated measures one-way ANOVA was performed between CIP and accuracy within participants. The ANOVA results were not significant, $F(1, 269)=2.60$, $p=.1081$. Residual analysis indicated that participant number two was an outlier. A Wilcoxon ranked sum test was performed and a significant difference between accuracy levels was detected, $x^2(1)=4.67$, p=.03. The results of the ANOVA were rejected in favor of the Wilcoxon test.

A final chart was created by plotting the variance of the CIP across identification blocks at each level of accuracy. There was a great deal of overlap between plot of variance at the 75% level (*M*=.016) and the plot of variance at the 90% level (*M*=.017). The chart is depicted in Figure 12.

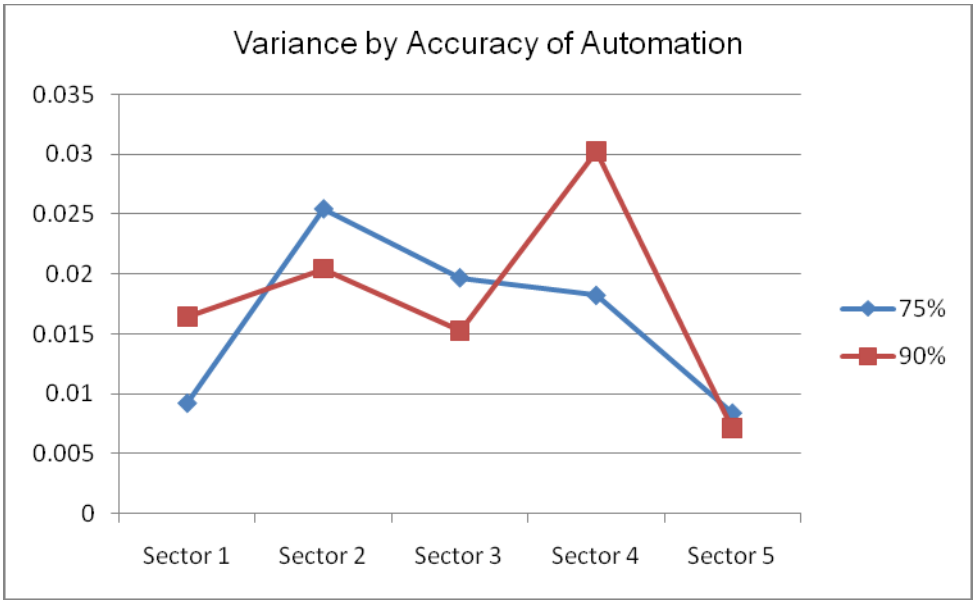


Figure 12. Variance By Accuracy of Automation

B. PERCIEVED RELIABILITY AND UTILITY

Participants were asked to estimate the reliability of the automated aid twice. One estimate was made at the conclusion of the 75% accuracy condition and the second estimate was made at the conclusion of the 90% accuracy condition. The mean estimated accuracy of the 75% condition was 81.4% (*SD*=8.7). The mean estimated accuracy of the 90% condition was 88.8% (*SD*=6.0). No correlation was found between the participant's total correct identification rate and the estimated accuracy of the automation, $r(58)=.05$, $p=.08$. A paired t-test was performed and the difference between accuracy estimates was found to be significant ($t(29)=6.11$, $p<.0001$).

Participants were also asked to indicate their preference for human or automation use in a future task. The question was asked at the end of the 75% and 90% accuracy trials. Participants were directed to choose between two responses: perform the task with no automated aid or allow the automated aid to perform the task without human interaction. The majority of participants (80%) preferred to perform the task with no automated aid. At the end of each trial, only three participants indicated a preference to allow the automation to completely control the task. Three additional participants misinterpreted the question and responded that they would prefer to perform the task and have the help of the automated identification system.

THIS PAGE INTENTIONALLY LEFT BLANK

V. DISCUSSION

A. HYPOTHESIS ONE

The hypothesis that increasing the level of automation would decrease the operator's ability to calibrate trust was partially supported. Trust calibration was measured by the percentage of targets correctly identified over five identification sectors. A difference in correct identification percentage between levels of automation was expected as an indication of trust calibration. A greater difference in CIP would indicate a greater difference in trust calibration. In the first identification sector, CIP was expected to be nearly equal. In subsequent sectors, the difference in CIP was expected to grow. Eventually, the CIP at all three levels of automation would converge. The result would be a plot of three different curves, beginning at roughly the same level of performance and ending at the same level of performance. A depiction of the expected results is presented in Figure 13. The actual line plot of results is presented in Figure 14.

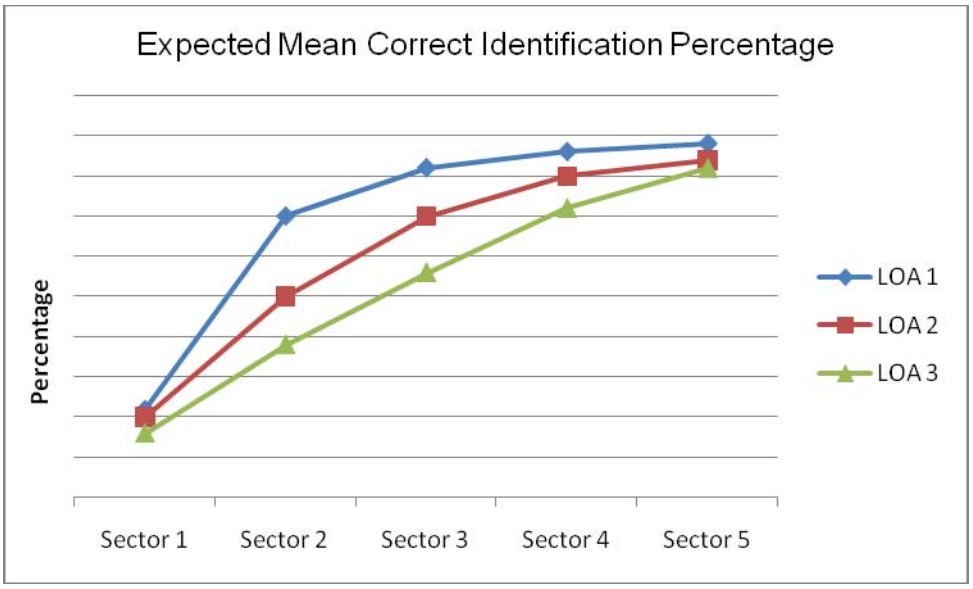


Figure 13. Expected Mean Correct Identification Percentage

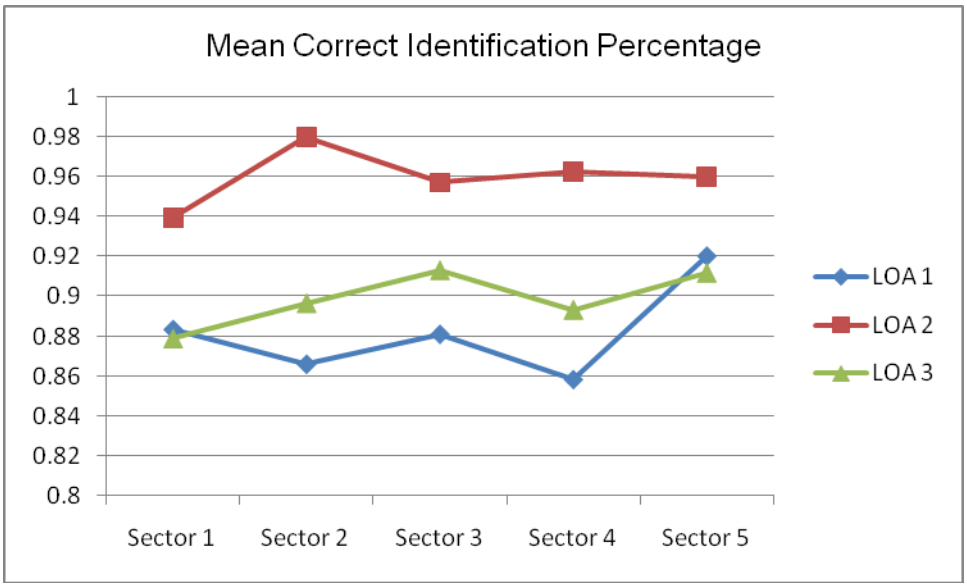


Figure 14. Actual Mean Correct Identification Percentage

A comparison between the plots of expected and actual results yielded noticeable differences. Correct identification percentage was expected to increase over time as an indication that trust calibration had increased, but the actual CIP remained relatively constant. In addition, plots of LOA 1 and LOA 3 overlapped on several occasions. Statistical analysis revealed that CIP at LOA 2 was greater than CIP at LOAs 1 and 3 ($p<.0001$). Results of the one-way ANOVA and Kruskal Wallis test were equal. The difference that was detected between levels of automation did not fully support the hypothesis. Against expectations, CIP at LOA 1 was significantly less than CIP at LOA 2. This may indicate that differing levels of trust calibration have occurred, but not in the expected manner.

A second indication of trust calibration was variance of the correct identification percentage. Smaller variance was expected to be an indication of better calibration with the automated system. We expected to see decreasing variance across sectors. In addition, as automation level increased, we expected to see higher variance and at each sector. A depiction of the expected results is presented in Figure 15. The actual variance plots is depicted in Figure 16.

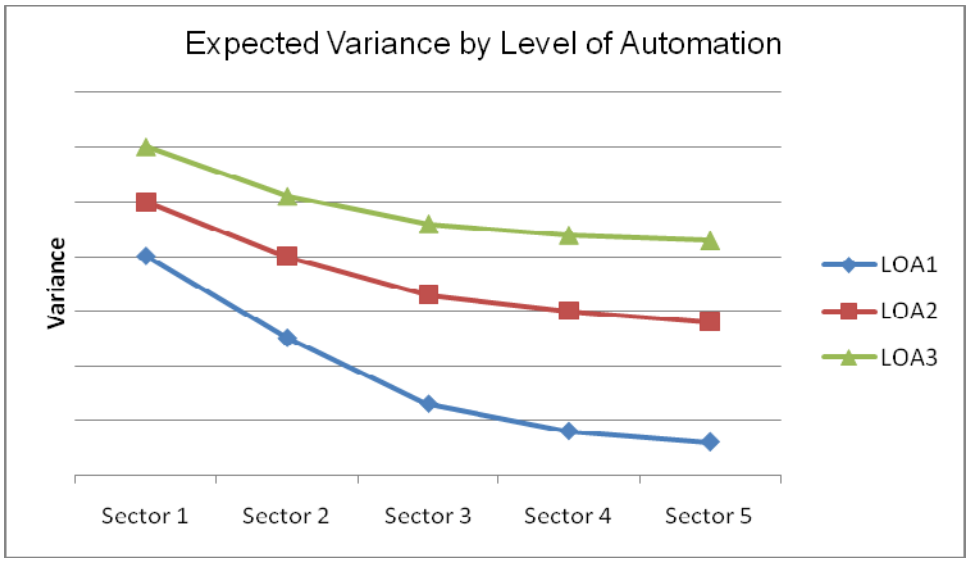


Figure 15. Expected Variance By Automation Level

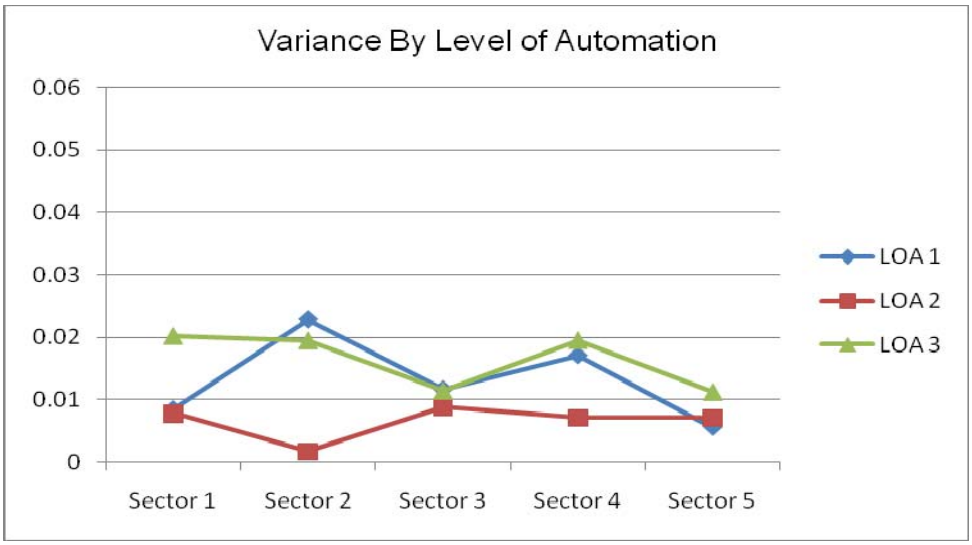


Figure 16. Actual Variance by Level of Automation

The expected results were not reflected in the actual variance plots. Though variance appeared to differ by LOA, there was no distinctive decrease in variance across identification sectors. The lack of change in variance across sectors indicates that trust calibration did not change over time. In addition, the similarity in variance by LOA indicates that calibration was the same at each level of automation.

The results can be viewed from three points of view. We will refer to Figure 17 to describe the possibilities. One explanation is that participants were able to calibrate their trust at LOA 2 more effectively. Correct identification percentage at LOA 2 was significantly higher than at the other LOAs. A higher correct identification percentage indicates better trust calibration. In other words, participants in LOA 2 had a better understanding of when to trust or distrust the automated system. Better calibration allowed them to recognize when the automated identification system accurately reflected ground truth.

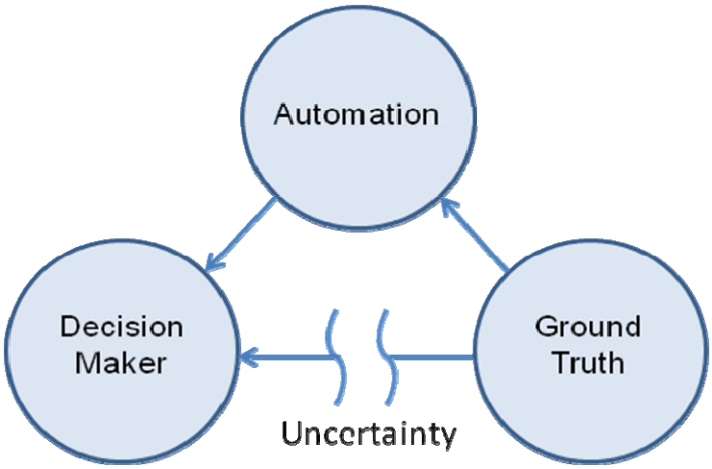


Figure 17. Relationship between ground truth, automation, and decision maker

Second, the lack of difference between LOA 1 and LOA 3. Sensitivity, the ratio of hits to false alarms, was very high. This suggests that participants may have been able to bypass the automated system and directly observe ground truth. Participants may have been able to make identifications based on ground truth with certainty.

Third, the high level of sensitivity exhibited by participants may suggest the task was too easy. If the task truly was too easy, it supports the idea that participants bypassed the automated system and directly interpreted ground truth. Several factors indicated that the identification task lacked difficulty. During the pilot experimental design, enemy and friendly targets possessed several distinguishing features. Participants in the first pilot study committed nearly zero

false alarms and misses. In the final experimental design, different enemy and friendly targets were chosen in an attempt to make the task more difficult. Even though the differences in enemy and friendly targets were reduced, participants quickly learned to correctly identify targets. Participants discovered visual cues that aided identification. In many cases, multiple targets were visible to participants at the same time. This allowed them to make direct comparisons between targets rather than absolute judgments. The software also made enemy and friendly targets point their weapons at each other, which aided identification. Finally, the participants' freedom to aim the video camera allowed them to dwell on many targets for extended lengths of time. The additional time enhanced identification of targets that may have been more difficult to distinguish. Any differences in correct identification percentage between levels of automation could be attributed to differences between the participant groups. Unfortunately, participants did not experience all three levels of automation, so this possibility could not be explored.

One shortfall of this study was that participants only experience one level of automation. The software in use did not allow participants to see the automated system at work. Their only knowledge of the automated system was provided in pretrial instructions. The instructions described how the automated system would function after an identification had been made. However, those functions were never performed in view of the participants. As a result, it would have been difficult to present the participants with all three levels of automation in a distinguishable manner.

Previous studies in human-automation interaction have attempted to measure trust in two ways. The first method was to directly question participants about their level of trust in the automation (Muir, 1989; Lee & Moray, 1992). These studies found that development of trust is a process that takes place over time. In addition, faults were shown to decrease trust in the system. In contrast with earlier work, the present study did not assess calibration of trust over time. Ruff, Narayanan, and Draper (2002) asked participants to report trust in systems

at three levels of automation. They found that participants exhibited the highest trust in the system when it operated at the management by consent level of automation. As the number of vehicles being controlled increased, reported trust levels also increased. This trend was not uniform across all levels of automation. In the management by exception condition trust decreased as number of vehicles increased. Performance was also greatest in the management by consent condition. The management by consent condition also led to the highest performance in the present study. This confirms the findings of Ruff, Narayanan, and Draper (2002). A second method to measure trust is to examine the decision to rely on the automation. Dzindolet et al. (2002) reported that, when asked to choose, participants indicated a reluctance to allow tasks to be fully controlled by automated systems. Participants in the present study confirmed those results by electing to perform tasks manually when given the choice between full manual control and full automation control.

The present study differed from previous research in that it attempted to measure the calibration of trust in a system over time. In previous studies, (van Dongen & van Maanen, 2006) calibration was indicated by a comparison between the estimated reliability in self and estimated reliability of automation. Under-trust in automated systems seemed to remain constant, but under-trust in self decreased from the first to second trial. Estimated trust was not collected in the present study, but participant performance did not change over time. This indicates that participant trust calibration did not change as trials progressed.

B. HYPOTHESIS TWO

The hypothesis that decreasing accuracy of the automation would decrease the operator's ability to calibrate trust was not supported. Trust calibration was measured by the percentage of targets correctly identified over five identification blocks. We expected to see a difference in correct identification percentage between levels of accuracy. In the first identification block, CIP was expected to be nearly equal. In subsequent blocks, the difference in CIP was

expected to grow. The CIP at the high accuracy level would always exceed the CIP at low accuracy level. The result would be a plot of two curves, beginning at roughly the same level of performance and ending at separate levels of performance. A depiction of the expected results is presented in Figure 18. The actual line plot of results is presented in Figure 19.

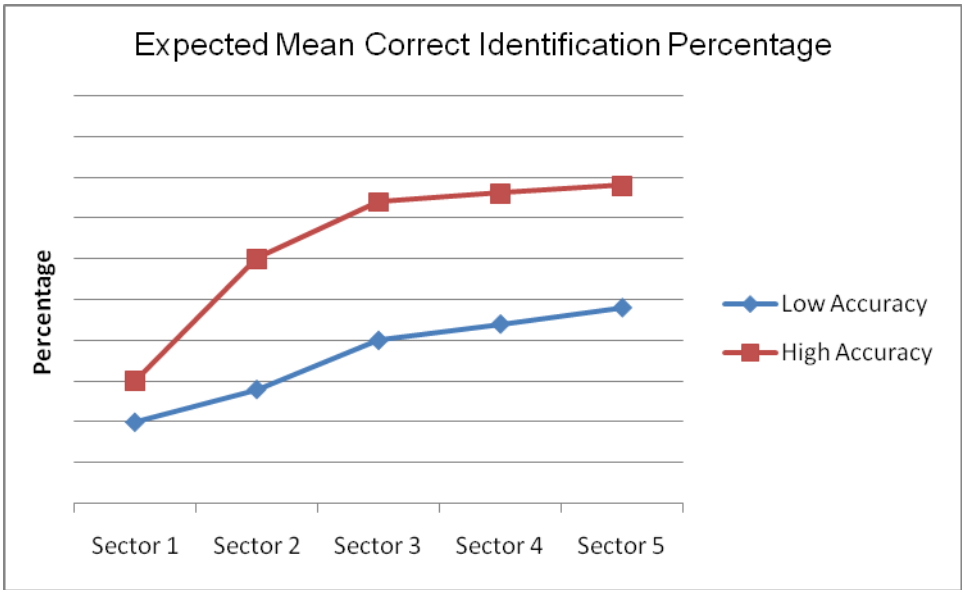


Figure 18. Expected Mean Correct Identification Percentage

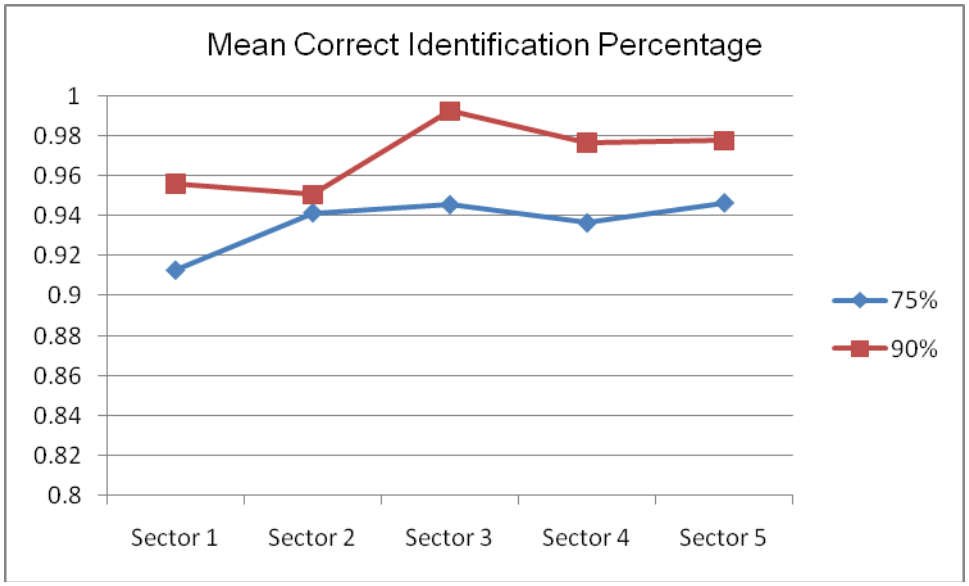


Figure 19. Actual Mean Correct Identification Percentage

The difference between CIP at 75% accuracy and 90% accuracy was found to be significant (p=.03). However, this finding does not support the hypothesis that trust calibration was greater at the higher level of accuracy. At the 75% accuracy level, participants were presented with 25 incorrectly identified targets. In the 75% condition, participants were able to identify 90.5% of targets correctly. However, 25% percent of the total targets were incorrectly identified by the automated system. This means that participants correctly recognized the error in the automated system 62% of the time. In the 90% accuracy condition, participants were able to identify 92.1% of the targets correctly. With 10% of the targets identified incorrectly by the automated system, the participants correctly recognized errors by the automated system 21% of the time. The difference in correct recognition of automated errors may indicate better trust calibration in the 75% condition.

A second indication of trust calibration was the variance of CIP across identification blocks. Smaller variance was an indication of increased calibration with the automated system. We expected to see decreasing variance across identification blocks. In addition, at the high accuracy level, we expected to see less variance at each identification sector. A depiction of the expected results is presented in Figure 20. The actual variance plot is depicted in Figure 21.

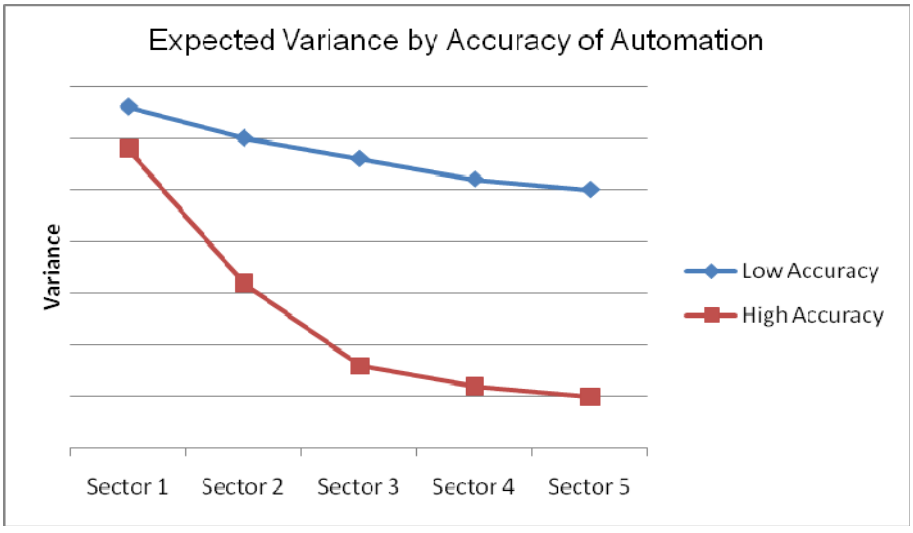


Figure 20. Expected Variance By Accuracy of Automation

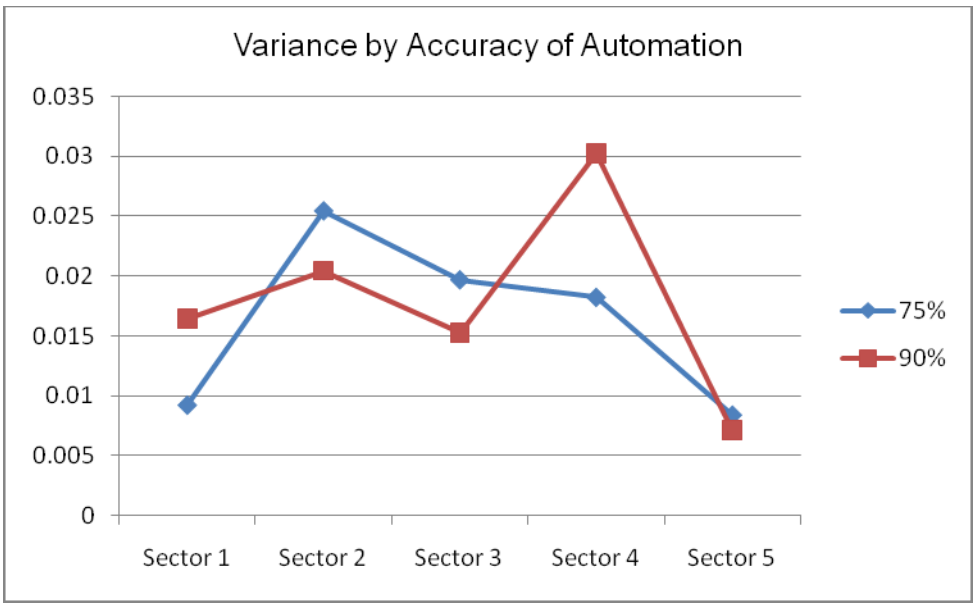


Figure 21. Actual Variance By Accuracy of Automation

The expected results were not reflected in the actual variance plots. The high accuracy variance plot consistently overlapped the low accuracy variance plot. In addition, there was no pattern of decreasing variance over time. Neither of the expected characteristics of the chart (decreasing variance and differing variance by accuracy) were present. The comparison of variance over time does not support the hypothesis that trust calibration will decrease with automation accuracy.

Participants were aware of differing accuracy levels. The matched pairs t-test of estimated accuracy indicated participants could distinguish the accuracy levels ($p<.0001$). This implies that participants knew when the automated system was incorrect and consistently disregarded the system's inputs. However, it does not support the hypothesis that trust calibration would be lower for less accurate systems.

Previous studies have demonstrated that changes in automation accuracy lead to changes in operator performance. Ruff, Narayanan, and Draper (2002) demonstrated that a decrease from 100% accuracy to 95% accuracy reduced operator efficiency from 76% to 69%. The present study confirmed the

relationship between accuracy and performance. However, trust calibration did not decrease with accuracy. Not all research has supported the results of Ruff, Narayanan, and Draper. Sorkin and Woods (1985) found that optimizing the performance of an automated system did not always yield the best results for the human-automation team.

Some have argued that levels of automation is not an effective way to describe human-automation interactions. The concept of “Levels of Automation” effectively explains the division of tasks between humans and automation, but it may not describe the way humans interact cognitively with automated systems. As Dekker and Woods (2002) point out, lists like Sheridan and Verplank’s (1978) Levels of Automation do not describe the cognitive processes that are involved in deciding how to use an automated system. It is possible that humans evaluate the trustworthiness of an automated system without considering its level of automation. In other words, after an error, trust in a system with high automation will change at the same rate as trust in a system with low automation. Humans may simply see an automated system that is making errors.

The present study does not support the claims of Dekker and Woods (2002). Significant differences in correct identification percentage by level of automation and accuracy were detected. These differences imply that LOA and accuracy affect the ability to calibrate trust in automated systems. However, differences in performance remained constant over time. Relatively constant performance may be explained by very high performance levels. The participants may have found the task to be too easy. If the task truly was too easy then participants may have disregarded the automated system or failed to use it as intended.

VI. CONCLUSIONS AND RECOMMENDATIONS

A. CONCLUSIONS

The employment of automated systems is expanding across the modern battlefield. The growth of automation has not eliminated the human from the system; it has transformed the human's role. With the trend toward increasing level of automation, humans have changed from operators to supervisors. This change does not necessarily mean that human workload has been reduced. Instead, cognitive resources are applied to different tasks, such as anticipating the automation and understanding the actions of the automation. Nonetheless, highly automated systems will be critical for the supervision of multiple unmanned systems across the battlefield.

The changing but continuous role that humans maintain with automated systems requires an understanding of the human-automation relationship. One aspect of this relationship that had not yet been fully explored was the process by which humans calibrate trust in automated systems. In the present study, participants were given the assistance of an automated system for the identification of enemy and friendly targets. Participants were placed in three groups and provided with information about the responsibilities of the automation. The descriptions corresponded to one of three levels of automation: decision support, management by consent, and management by exception. In addition, the participants experienced two levels of automation accuracy, 75% and 90%.

Two hypotheses were proposed. The first hypothesis, that a high level of automation would decrease the ability to calibrate trust was partially supported. In addition, the second hypothesis, that low automation accuracy would decrease the ability to calibrate trust was not supported.

The results of this study suggest that a system's level of automation may influence an operator's ability to calibrate trust. Participants who had been told they were using automation that employed management by consent

outperformed those using decision support and management by exception levels of automation. Better performance may indicate better trust calibration, but not in the direction hypothesized. The accuracy of the automated system also influenced the correct identification percentage. Performance was better at the 90% accuracy level but a greater percentage of automation errors were identified at the 75% accuracy level. We hypothesized that trust calibration would decrease as accuracy decreased, but trust calibration appeared to increase as accuracy decreased.

The difference in performance between levels of automation could also be explained by the experimental design. Subjects only experienced a single level of automation, so we don't know if their performance would be equal across levels. An additional limitation of the experiment was the ease of distinguishing enemy and friendly personnel. Participants were better at identifying targets than the automated system. As a result, participants may not have used the automated system as expected. Several participants stated that the automation was helpful in finding the location of targets, but the friend or foe indications were ignored because the participant may have believed he or she was more accurate.

B. RECOMMENDATIONS FOR FOLLOW-ON RESEARCH

Research in the area of trust and levels of automation is relatively new. There are many opportunities to expand our understanding of the human-automation relationship. Future research into levels of automation would certainly benefit from greater automation functionality. An automated system with more fidelity, that allows participants to see the outcome of their decisions may create a more noticeable difference in the levels of automation experienced. It would also allow a repeated measures design for more effective statistical analysis. Future experiments should be designed with a more difficult task for participants. Measurement of trust and trust calibration would be more effective if the participants actually needed to rely on the automated system.

In the course of the present study, some additional questions were developed. Previous studies found that humans often choose not to rely on automated systems (Dzindolet et al., 2002). One direction for future research would be to focus on the human's decision to rely on automated systems. At what degree of difference between human and automation performance will humans prefer to rely on the automated system? Will that preference vary by level of automation?

Another direction for research should look at alternate methods for measuring trust calibration. The present study equated percent of targets correctly identified to proper calibration. However, many participants reported using the automated system in a manner other than intended. These participants used the automated system to point to the location of targets, but ignored the target type indicated by the automated system. Since the participants knew to trust their own judgment, one might assume they were properly calibrated all along. This implies that a better indication of trust calibration may be the participant's knowledge of when to trust and when to distrust the automated system. Many automated systems have reliabilities that change as the conditions and environment change. A human that knows when to trust and when to distrust the automated system is highly calibrated. Simply achieving a high number of correct answers may not indicate high calibration.

An understanding of the cognitive processes in play on a human-automation team is vital to the future integration of highly automated systems onto the battlefield. We must examine the manner in which humans build trust in automated systems and how trust relates to effective operation. The real goal of future research should not be to divide the tasks between humans and machines; the efforts need to focus on how humans and machines work together.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. POST-TRIAL QUESTIONNAIRES

Post-Trial 2 Questions

In the previous scenario you were presented with 100 targets. Please estimate the percentage of times the automation was correct in its identification of individuals. _____

If you were to perform this task in the future would you prefer to execute the mission unaided or with the use of the automated identification system and why?

Additional Comments:

Post-Trial 3 Questions

In the previous scenario you were presented with 100 targets. Please estimate the percentage of times the automation was correct in its identification of individuals. _____

If you were to perform this task in the future would you prefer to execute the mission unaided or with the use of the automated identification system and why?

Additional Comments:

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. DEMOGRAPHIC QUESTIONNAIRE

This experiment is designed to explore what people think about automation. By automation, we mean devices and systems that make work or other tasks easier for you to do. Examples of automation are: Global Positioning System (GPS) receivers you might have in your car, handheld electronic organizers, or military applications such as integrated data displays or threat warning systems.

1. Gender_____ 2. Age_____ 3. Branch of Service_____ 4. Time in Service_____

5. Are you color blind? Yes / No

6. Think about your last military job before arriving at NPS. In that job, how often did you use automated devices?

   ___Daily ___Weekly ___Once a Month

   ___Several Times a Year ___About Once a Year ___Less Than Once A Year

7. Now that you are an NPS student, how often do you use automated devices?

   ___Daily ___Weekly ___Once a Month

   ___Several Times a Year ___About Once a Year ___Less Than Once A Year

8. Please indicate how comfortable you are with the automation you have used:

	I prefer to never use automation				I prefer to use automation whenever possible	
In you military job	☐	☐	☐	☐	☐	☐
In your time at NPS	☐	☐	☐	☐	☐	☐

THIS PAGE INTENTIONALLY LEFT BLANK

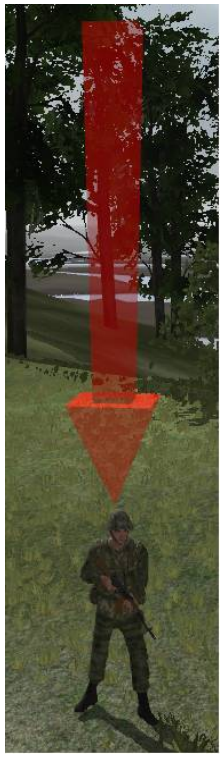
APPENDIX C. EXPERIMENTAL INSTRUCTIONS

INITIAL INSRUCTIONS

There has been a recent increase in terrorist activities along a road of strategic importance. Little is known about the position of enemy and friendly personnel along the roadway. In order to collect intelligence on the disposition of forces an unmanned aircraft has been directed to scout the roadway. You are the operator of the unmanned aircraft video feed. The aircraft will fly a programmed route over the roadway while you manipulate the camera. When you encounter an enemy along the route press the key labeled E to indicate an enemy, the target will be destroyed when the aircraft has passed. When you encounter friendlies along the route, press the key labeled “F” to keep them safe from engagement.

Decision Support—Trial 1

Your unmanned aircraft has been tasked to scout another roadway for enemy and friendly personnel. Since your last mission the sensor package has been upgraded with an automated identification system. The system compiles visual indications and additional sensor data then indicates the possible presence of enemy or friendly units. As the operator, you must evaluate the video feed and the indications of the automated identification system to mark enemy and friendly personnel. Your own inputs, using the “E” (enemy) and “F” (friendly) keys will ultimately determine which units are to be engaged.

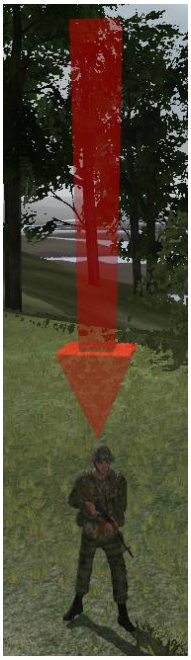


Decision Support—Trial 2

Your unmanned aircraft has been tasked to scout a third roadway for enemy and friendly personnel. As the operator, you must once again evaluate the video feed and the indications of the automated identification system to mark enemy and friendly personnel. Your own inputs, using the “E” (enemy) and “F” (friendly) keys will ultimately determine which units are to be engaged.

Management by Consent—Trial 1

Your unmanned aircraft has been tasked to scout another roadway for enemy and friendly personnel. Since your last mission the sensor package has been upgraded with an automated identification system. The system compiles visual indications and additional sensor data, and then determines if a unit is enemy or friendly. Identified personnel will be indicated on the video feed. The system will autonomously direct the engagement of enemy targets and mark friendly units for safety. As the operator, you must provide consent before the automated system will transmit target information. Evaluate the video feed and the indications of the automated identification system to make your determination and provide consent by using the “E” (enemy) and “F” (friendly) keys. When you disagree with the automation please mark the target in the appropriate manner. If you do not mark a target or provide consent, no enemy or friendly information will be passed.

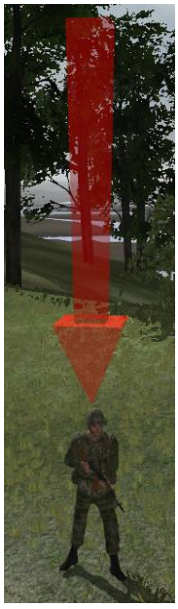


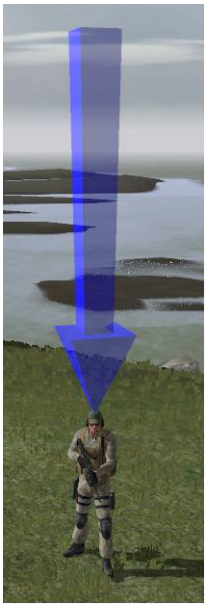
Management by Consent—Trial 2

Your unmanned aircraft has been tasked to scout a third roadway for enemy and friendly personnel. The system will autonomously direct the engagement of enemy targets and mark friendly units for safety. As the operator, you must provide consent before the automated system will transmit target information. Evaluate the video feed and the indications of the automated identification system to make your determination and provide consent by using the “E” (enemy) and “F” (friendly) keys. When you disagree with the automation please mark the target in the appropriate manner. If you do not mark a target or provide consent, no enemy or friendly information will be passed.

Management by Exception—Trial 1

Your unmanned aircraft has been tasked to scout another roadway for enemy and friendly personnel. Since your last mission the sensor package has been upgraded with an automated identification system. The system compiles visual indications and additional sensor data, and then determines if a unit is enemy or friendly. Identified personnel will be indicated on the video feed. The system will autonomously direct the engagement of enemy targets and mark friendly units for safety. As the operator, you have the ability to veto the automated system. Without a veto, the targeting information will be transmitted. Evaluate the video feed and the indications of the automated identification system to make your determination. You may veto the automation by pressing the “D” (disagree) key and agree with the automation by pressing the “A” (agree) key.





Management by Exception—Trial 2

Your unmanned aircraft has been tasked to scout a third roadway for enemy and friendly personnel. Since your last mission the sensor package has been upgraded with an automated identification system. The system compiles visual indications and additional sensor data, and then determines if a unit is enemy or friendly. Identified personnel will be indicated on the video feed. The system will autonomously direct the engagement of enemy targets and mark friendly units for safety. As the operator, you have the ability to veto the automated system. Without a veto, the targeting information will be transmitted. Evaluate the video feed and the indications of the automated identification system to make your determination. You may veto the automation by pressing the “D” (disagree) key and agree with the automation by pressing the “A” (agree) key.

LIST OF REFERENCES

32nd Army Air and Missile Defense Command (2003). *Patriot Missile Defense Operations During Operation Iraqi Freedom*. Washington, DC: United States Army.

Dekker, S. W., & Woods, D. D. (2002). MABA-MABA or abracadabra? Progress on human-automation coordination. *Cognition, Technology, and Work*, *4*, 240–244.

Defense Acquisition University (2010). Retrieved March 11, 2011, from http://icatalog.dau.mil/onlinecatalog/courses.aspx?crs_id=1726

Department of Defense. (2007). *Unmanned Systems Roadmap 2007-2032*. (AD No. ADA475002). Washington, DC: Defense Technical Information Center.

Department of the Air Force. (2009a). *United States Air Force Unmanned Systems Flight Plan 2009-2047*. (AD No ADA505168). Washington, DC: Defense Technical Information Center.

Department of the Air Force. (2009b). *United States Air Force Human Systems Integration Handbook*. (AD No AFD-090121-054). Washington, DC: Defense Technical Information Center.

Barber, B. (1983). *The logic and limits of trust*. New Brunswick, NJ: Rutgers University Press.

Dongen, K. Van, & Maanen, P. P., Van (2006). Under-reliance on the decision aid: A difference in calibration and attribution between self and aid. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 339–343). Santa Monica, CA: Human Factors and Ergonomics Society.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (1999). The misuse and disuse of automated aids. In *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (pp. 339–343). Santa Monica, CA: Human Factors and Ergonomics Society.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2001). *A framework of automation use* (Technical Report ARL-TR-2412). Aberdeen Proving Ground, MD: Army Research Laboratory.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2002). The perceived utility of human and automated aids in a visual detection task. *Human Factors*, 44, 79-94.

Endsley, M. (1987). The application of human factors to the development of expert systems for advanced cockpits. In *Proceedings of the Human Factors and Ergonomics Society 31st Annual Meeting* (pp. 1388-1392). Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42, 462–492.

Fisher, C. W., & Kingma, B. R. (2001). Criticality of data quality as exemplified in two disasters. *Information and Management*, 39, 109–116.

Fitts, P. M. (1951). *Human Engineering for an Effective Air Navigation and Traffic Control System*. Washington, DC: National Research Council.

Guerlain, S. (1995). Using the critiquing approach to cope with brittle expert systems. In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (pp. 233–237). Santa Monica, CA: Human Factors and Ergonomics Society.

Guerlain, S., & Bullemer, P. (1996). User-initiated notification: A concept for aiding the monitoring of process control of operators. In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 283-287). Santa Monica, CA: Human Factors and Ergonomics Society.

Lee, J. D., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35, 1243–1270.

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, 40, 153–184.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46, 50–80.

Masalonis, A. J., & Parasuraman, R. (1999). Trust as a construct for evaluation of automated aids: Past and future theory and research. *Proceedings of the Human Factors and Ergonomics Society*, 43rd Annual Meeting, 184–188. Santa Monica CA: Human Factors and Ergonomics Society.

Mosier, K. L., & Skitka, L. J. (1996). Human decision makers and automated decision aids: made for each other? In R. Parasuraman & M. Mouloua (Eds., *Automation and human performance: Theory and applications* (pp. 201–220). Hillsdale, NJ: Erlbaum.

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, 27, 527–539.

Muir, B. M. (1989). Operators' trust in and use of automatic controllers in a supervisory process control task. Unpublished doctoral dissertation, University of Toronto, Toronto, Canada.

Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, 37, 1905–1922.

Mullen, B., Johnson, C., & Salas, E. (1991). Productivity loss in brainstorming groups: A meta-analytic integration. *Basic and Applied Psychology* 12, 3–23.

Naval Postgraduate School. (2010). Retrieved March 11, 2011, from https://cle.nps.edu/xsl-portal/site/cb062e07-d8d7-40a9-955f-c40b1e9098d7/page/dc9d196d-854b-4b0d-a8ed-68c003b14a80

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, and abuse. *Human Factors*, 39, 230–253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, man and Cybertnetics-Part A: Systems and Humans*, 30, 286–297.

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology*, 49, 95–112.

Rouse, W. B., & Rouse, S. H., (1983). *A framework for research on adaptive decision aids.* (Technical Report AFAMRL-TR-83-082). Wright-Patterson Air Force Base, OH: Air Force Aerospace Medical Research Laboratory.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *The Academy of Management Review*. 23, 393–404.

Ruff, H. A., Narayanan, S., & Draper, M. H. (2002). Human interaction with levels of automation and decision-aid fidelity in the supervisory control of multiple simulate unmanned air vehicles. *Presence: Teleoperators and Virtual Environments*, 11, 335–351.

Sheridan, T. B. (2002). *Humans and automation: System design and research issues*. Santa Monica, CA: Human Factors and Ergonomics Society.

Sheridan, T. B., & Verplank, W. L. (1978). *Human and Computer Control of Undersea Teleoperators*. Cambridge, MA: Man-Machine Systems Laboratory, Department of Mechanical Engineering, MIT.

Sorkin, R. D., & Woods, D. D. (1985). Systems with human monitors: A signal detection analysis. *Human-Computer Interaction*, 1, 49–75.

Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychology Review*, 61, 1954.

Wickens, C. D., Lee, J. D., Liu, Y., & Gordon-Becker, S. E. (2004). *An Introduction to Human Factors Engineering*. Upper Saddle River, NJ: Pearson Education.

Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on trust and reliance. *Theoretical Issues in Ergonomic Science*, 2, 352–367.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Lawrence G. Shattuck
   Naval Postgraduate School
   Monterey, California

4. LTC Robert L. Shearer
   Naval Postgraduate School
   Monterey, California